



University of Pennsylvania  
**ScholarlyCommons**

---

Dropsie College Theses

Dropsie College for Hebrew and Cognate  
Learning

---

Spring 3-11-1965

## Evaluation in Jewish Education in the United States with Special Emphasis on Testing, 1910 - 1965

A. Hillel Henkin

Follow this and additional works at: <https://repository.upenn.edu/dropsietheses>

 Part of the [Bilingual, Multilingual, and Multicultural Education Commons](#), [Curriculum and Instruction Commons](#), [Educational Assessment, Evaluation, and Research Commons](#), [Educational Methods Commons](#), [History Commons](#), [Jewish Studies Commons](#), and the [Language and Literacy Education Commons](#)

---

Library at the Katz Center - Archives Thesis. LC719 .H465 1965.

This paper is posted at ScholarlyCommons. <https://repository.upenn.edu/dropsietheses/115>  
For more information, please contact [repository@pobox.upenn.edu](mailto:repository@pobox.upenn.edu).

---

## Evaluation in Jewish Education in the United States with Special Emphasis on Testing, 1910 - 1965

### Abstract

This study aims to review the techniques used in testing and evaluating Jewish education, to analyze their inadequacies and to indicate how Jewish education may benefit from the best accepted standards in general evaluation. It is written with a conviction and with a specific objective in mind. The conviction is the vital importance of evaluation in education in general and specifically in Jewish education. The objective is to contribute toward creating the instruments for the proper implementation of evaluation and the widespread use of these instruments.

### Degree Type

Dissertation

### Degree Name

Doctor of Philosophy (PhD)

### First Advisor

William Chomsky

### Second Advisor

T.E. McMullin

### Third Advisor

Meir Ben-Horin

### Subject Categories

Bilingual, Multilingual, and Multicultural Education | Curriculum and Instruction | Educational Assessment, Evaluation, and Research | Educational Methods | History | Jewish Studies | Language and Literacy Education

### Comments

Library at the Katz Center - Archives Thesis. LC719 .H465 1965.

---



EVALUATION IN JEWISH EDUCATION

IN THE UNITED STATES

WITH SPECIAL EMPHASIS

ON TESTING

1910 - 1965

by

A. Hillel Henkin

A dissertation

submitted in partial fulfillment  
of the requirements for the degree of  
Doctor of Education  
in the School of Education

The Dropsie College  
for Hebrew and Cognate Learning  
Philadelphia  
1965

APPROVAL

This dissertation, entitled  
Evaluation in Jewish Education  
in the United States  
With Special Emphasis on Testing,  
1910-1965

by

A. Hillel Henkin

Candidate for the degree of  
Doctor of Philosophy

has been read and approved by

William Cronin  
T. E. McMillin  
Max R. R. R.

Date March 11, 1965



## TABLE OF CONTENTS

I.	INTRODUCTION	1
II.	EVALUATION IN GENERAL EDUCATION	
1.	Principles of Evaluation	
a)	Definition and Scope	7
b)	Purpose	10
c)	Tests, Measurement and Evaluation	11
d)	Problems of Evaluation	13
e)	Steps in Evaluation	15
2.	Standards of Evaluation	
a)	Validity	16
b)	Reliability	20
c)	Objectivity	23
d)	Norms	24
e)	Practicability	27
f)	Summary of Standards of Test Evaluation	28
3.	Instruments of Evaluation	30
a)	Objective Tests	31
b)	Non-Objective Tests	41
c)	Observation	43
d)	Anecdotal Records	45
e)	The Questionnaire	47
f)	Check Lists and Rating Scale	50
g)	Interview	53
h)	Projective Techniques	55
i)	Cumulative Records	56
j)	The Case Study	58
III.	APPLICABLE TECHNIQUES IN EVALUATING JEWISH EDUCATION	
1.	The All-Day School	60
2.	The Supplementary Jewish School	61
3.	The Aims of the Jewish School	62
4.	Learning	66
a)	Intelligence Testing	66
b)	Aptitude Testing	67
c)	Achievement Tests	68
(1)	Divisions in the Jewish School	68
(2)	Subjects of Study-General	69
(3)	Subjects of Study on Elementary Level	70
(4)	Feasible Tests on Elementary Level	73
(5)	Subjects of Study on Secondary Level	76
5.	Practices	79
6.	Values and Attitudes	81



#### IV. SURVEY AND ANALYSIS OF EVALUATIVE INSTRUMENTS IN JEWISH EDUCATION

##### 1. Historical Review

a) Publications	83
b) National Organizations	85
c) Central Agencies	87
d) School Testing	95
e) Classroom tests	100

##### 2. Intelligence Tests 103

##### 3. Aptitude Tests 106

##### 4. Achievement Tests

###### Introduction 110

a) Hebrew	112
b) Siddur	125
c) Holidays and Observances	128
d) Bible	134
e) History	140
f) Miscellaneous	147

##### 5. Practices 149

##### 6. Values and Attitudes 155

a) Ethical Behavior	155
b) Attitude Toward Religious School	160
c) Jewish Values	165
d) Theological Concepts	168

##### 7. Summary 170

#### V. RECOMMENDATIONS 173

#### VI. BIBLIOGRAPHY 176

#### VII. APPENDICES

a) Questionnaires Used In Survey	
b) Selected Tests	
c) Comparative List of Aims of Jewish Education	
d) Standard Deviation, Spearman-Brown Formula and Kuder-Richardson Formula #21	



## TABLES

	<u>Page</u>
1. Subjects Listed in Courses of Study for Elementary Grades	72
2. Subjects Taught in Seven Hebrew High Schools	77
3. Review of Testing in Central Agencies	91
4. Agency Testing According to Subjects	93
5. Agency Testing According to Grades	94
6. Non-Agency Testing According to Grades	97
7. Comparison of Degree of Testing in Agency and Non-Agency Schools	98
8a. Analysis of Validity, Reliability and Norms in Selected Tests in Hebrew	120
8b. Analysis of Objectivity and Practicability in Selected Tests in Hebrew	121
9a. Analysis of Validity, Reliability and Norms in Selected Tests in Holidays and Observances	132
9b. Analysis of Objectivity and Practicability in Selected Tests in Holidays and Observances	133
10a. Analysis of Validity, Reliability and Norms in Selected Tests in Humash	138
10b. Analysis of Objectivity and Practicability in Selected Tests in Humash	139
11a. Analysis of Validity, Reliability and Norms in Selected Tests in Jewish History	143
11b. Analysis of Objectivity and Practicability in Selected Tests in Jewish History	144
12a. Analysis of Validity, Reliability and Norms in Available Bar Mitzvah Tests	148
12b. Analysis of Objectivity and Practicability in Available Bar Mitzvah Tests	148



## INTRODUCTION

This study aims to review the techniques used in testing and evaluating Jewish education, to analyze their inadequacies and to indicate how Jewish education may benefit from the best accepted standards in general evaluation. It is written with a conviction and with a specific objective in mind. The conviction is the vital importance of evaluation in education in general and specifically in Jewish education. The objective is to contribute toward creating the instruments for the proper implementation of evaluation and the widespread use of these instruments.

Jewish education in America suffers from many handicaps: The absence of a Jewish home, an environment conducive to assimilation with the majority culture, the lack of time. Furthermore the child comes to the Hebrew school tired from a full-day of study at the public school. The average Jewish child also starts his religious education at a late age and ends it too soon, normally after only four or five years of study.

These almost insuperable obstacles require herculean efforts to overcome them. They require strong, positive and farsighted leadership and excellent and dedicated teachers. The latter, however, are almost unavailable because the teaching profession has yet to attain a satisfactory economic and social status. This results in part from the part-time nature of Hebrew teaching and the small numbers attending secondary Jewish schools from which must come the reservoir for Hebrew teacher colleges. The need generally to attend both a Hebrew teacher's college and a secular institution of higher learning also greatly decreases the numbers entering the teaching profession.

A major disturbing factor has been the virtual boycott over a long period of time of any attempt on the part of Jewish educators to establish objective evaluative techniques. In some cases this is due to a lack of knowledge of



testing techniques and the absence of effective training in the area. This may, however, also be due to a subconscious fear of the results of such testing. One educator says, "Past experience indicates that standard tests do not fairly evaluate progress based on school program." One says he never found a program to meet his needs. One frankly states he doesn't test because he doesn't "feel the need for it." One educator places the major blame upon a general inertia which pervades Jewish education and gives the following additional causes: feared loss of time, possible cost, overburdened administrators, lack of personnell for proper interpretation and lack of facilities to apply the results of evaluation. Varied factors are undoubtedly the cause of the total negation of the concept of objective testing- almost universally practiced in general education- by a noted Jewish educator, as expressed in the following words:

הבחינות שבמחלקה זו, הן כמעט תמיד, אינן אלא אמצעים  
והן מיועדות, שכן מחלקות אלה הן מחלקות  
כולן בקיאות, הן הן "מחלקות" הן הן  
ואם יש להן, הן הן, הן הן, הן הן, הן הן  
הן הן, הן הן, הן הן, הן הן, הן הן, הן הן  
(62, pp. 59-60) "המחלקה"

It must be granted that testing, including objective testing, is not

foolproof and is open to doubt. As one author puts it, "Even though both their actual construction and their ability to measure have been the subject of very recent attack, it cannot be too strongly emphasized that judicious use of tests can be an invaluable aid to education. It can give teachers and administrators insight into the areas in which a child excells, and those in which he is weak. Testing results are able to indicate broadly the need for more emphasis in a particular area of the school curriculum. Guidance conferences can be made more meaningful and fruitful when it is remembered that achievement tests are constructed as individual tests and the scores are utilized during these individual



conferences. On the other hand, a well-administered testing program clearly emphasizes the fact that a test score must be considered only a well-calculated estimation, since a statistical figure is not an absolute number." (93)

True, the need for testing has been strongly espoused by some Jewish educators. J. B. Maller long ago urged the introduction of standardized objective tests which are "one of the outstanding characteristics of the modern secular school...Only with the aid of such tests can the efficiency of schools be measured or improved or the work of different schools or school systems be measured." (53) Zdanek Vanek said, "In disregarding testing techniques educators deny themselves the only valid method for evaluating their own effectiveness." (46) Edward Nedelman wrote, "Accurate measurement in the field of Jewish education has long constituted a real need." (52)

Do we really know the results of the Jewish education of our children? How <sup>is</sup> much achievement testing/conducted in Jewish education, what is its nature and how valid are these tests? Testing is usually conducted by the individual teacher. Are they objective or do they merely aim to convince the teacher himself that he's succeeding in his teaching? One or two illustrations may suffice. In a Hebrew class the children often read a story, try to translate it with the help of the vocabulary on the bottom of the page or in their notebooks and when they find difficulties the teacher helps them. They thus review it several times and the teacher feels they know it. He then gives them a test by writing on the board a number of sentences which they are to translate. Perhaps many pupils obtain good marks on such tests but this may be so because they recognize words in relation to others in the sentence. Can they recognize the vocabulary in a different context, can they understand different stories with the same vocabulary? If not, what have they learned? Even when there is recognition, this is actually the lowest form of knowledge. Can they form sentences or answer questions in the language and otherwise use the skills acquired?



Let us assume, however, that various techniques are used to have the student absorb the material. Does it mean that he will remember it a month or a year later, that it will truly become part of him? How many test the degree of retention after the passage of time? How much do we know if one method or another lends itself to better retention for a particular group or generally?

There is an ongoing controversy over homogeneous vs. heterogeneous grouping. Is the problem the same in Jewish as in general education? How well are we trying to determine the atypical children so that they may be cared for properly? And do we test aptitude to determine in advance satisfactory groupings and subjects of study?

Achievement tests measure knowledge, information. Do we test attitudes and the effect of Jewish education in molding or modifying them? For we generally agree that the basic value of Jewish education is not to know answers to questions someone may ask or to be able to discuss them intelligently or to convey them to the next generation, but rather the molding of character, instilling within the individual those characteristics which are exemplified in our great men, which gave strength and character to our people and made them great. In training our children to practice Judaism and to follow the paths of our fathers we endeavor to instill in them the ideals for which Judaism stood. Are we succeeding in this direction? Are we making an effort to find out?

In evaluating both achievements and attitudes a basic question is: Does the class, the institution, the system have a clear set of objectives? We cannot measure progress unless we know what we are trying to accomplish. The objectives are not synonymous with the course of studies, although the course of studies may help realize specific objectives and we may discover them by a careful review of the course. We are here concerned not with what these objectives are but rather with their existence,



their relation to the course of studies and the degree of their realization. What are our objectives in teaching Hebrew, history, laws and customs? Knowledge of the text, it may be said. However, there are different types and degrees of knowledge. Does it include speaking the Hebrew of the text, knowing the grammar, ability to spell; does it include practicing the laws, does it include remembering dates and biographies- does it include becoming imbued with the spirit of the studies? Furthermore, what are our objectives beyond information? Are we trying to instill a desire to live an ethical life, a feeling of unity with *Time* *He*, a loyalty to Torah, devotion to the land of Israel, a readiness to sacrifice for Jewish values? If we are, are the texts geared in these directions, are the teachers aware of these objectives, do they believe in them and are they endeavoring to instill them? There may be other objectives or additional ones. Whether immediate or long-range they must be spelled out clearly. No evaluation is possible unless the program has specific objectives.

Finally, where there are objectives what is their relation to the course of studies, with the means of their realization? If our objective is the ability to participate in a religious service, is enough time devoted to enable its accomplishment? Related to this is how much time is generally necessary to gain this objective? If one objective is to instill within the children a love for the land of Israel, does the program actually provide sufficient opportunity to realize it, both insofar as teachers and course are concerned? How much is the aim to instill and enhance ethical values being realized? Of course, these are difficult to measure, partly because they involve extraneous influences, partly because so many different factors are involved, and finally because completely valid instruments may never be created- but is an effort being made in that direction?



We continuously say that the teacher is the basic, indispensable ingredient in conveying learning. How much are we trying to determine what makes a good teacher- is it his background, his training, his experience, his personality? How much are we trying to determine the success of the teacher and how one teacher or type of teacher is preferable to another? Educators stress more learning, higher degrees as making better teachers. Is this actually so?

In short, are we leaving Jewish education to change, hoping that everything will somehow be resolved satisfactorily, or are we trying to find out how successful it is and how it can be improved? There always is some form of testing but it is usually subjective and short-range, haphazard and planless. Do we have the necessary machinery for evaluation and are we using it satisfactorily? If not, what do we need, how do we obtain it and how do we use it?

Finally, what benefits would be derived from an adequate procedure of evaluation in Jewish education? Surely we should want to know if our work is successful. Knowing that we have failed may be a first step toward improving conditions. We may never know nor agree on all factors making for success in so complicated an effort as Jewish education but we must work toward that goal. A good system of evaluation may spell the difference between success and failure in many areas of learning and teaching.



## II. Evaluation in General Education

### 1. Principles of Evaluation

#### a) Definition and Scope

Evaluation is "a comprehensive, cooperatively-developed, continuous process of inquiry which, in the final analysis, must be interpreted and defined in terms of its principles, functions, characteristics and purposes."

(1) It signifies "describing something in terms of selected attributes and judging the degree of acceptability or suitability of that which has been described." (2) It "constitutes a process of appraising the progress of pupils toward the values or objectives of the curriculum." (3) It is "fundamentally only the process by which we find out how far the objectives of the institution are being realized." (4)

The definition of evaluation thus involves the following factors:

(1) It is comprehensive a) in objectives: It deals with all major objectives of the school or school system, with qualitative and quantitative- not subject matter achievement only. It is concerned with pupil behavior and the realization of values. b) In means: It uses a variety of means of appraising achievement, attitude, personality and character. c) In results: It endeavors to create a full picture of the condition of the individual, the educational situation or the institution, and the effectiveness of the course of studies in carrying out the objectives. (5)

This presupposes a) The existence of a curriculum and of clear objectives which the curriculum is endeavoring to realize. The objectives are drawn from three areas: (1) structure, plant, equipment or organization, (2) process-aspects of school or class procedure, quality of the program, the teaching,



the instructional materials. (3) product- performance, indicating ability, interest, attitude or adjustment. b) Since it is impossible to evaluate everything it clearly delineates what attributes it is endeavoring to appraise.

(2) It is a cooperative endeavor- to be most effective it must include the view and the efforts of different people connected with and well-versed in the educational endeavor. It must also relate these various reactions.

(3) It is continuous- it is concerned with an ongoing educational process, not a completed one. In order to serve its purpose best it must be followed up from time to time. In a learning situation it should take place at the beginning, during and at the end of the experience.<sup>(40)</sup> At the beginning it helps to clarify purposes and discover needs, in the middle to devise means of overcoming difficulties, at the end to measure accomplishments and reevaluate objectives.

Wrightstone (6) sets the following additional hypotheses of evaluation.

(1) Curriculum change and evaluation are aspects of true experimentation. Both result in providing more insight and richness for the reconstruction of pupil personalities.

(2) For a variety of major objectives no adequate methods of collecting reliable evidence are available; until these are evolved, they must be appraised by as careful subjective means as possible.

(3) The measures will correspond to functional units of pupil behavior in actual curricular situations rather than in isolated units.

(4) Reliable and valid objective instruments of measurement are restricted to limited aspects of pupil behavior. It is impossible to measure the whole



result of the educative experience by any one test or battery of tests. It is hoped that by measuring many important and vital aspects some valid appraisals may be made.

(5) Another hypothesis, accepted by many educators, is that measuring functional behavior can best be developed by teachers working in cooperation with test technicians.

Boykin (1,7) adds the following principles:

- (1) There must be a clearcut concept of the aims of society and of education.
- (2) Self-analysis and appraisal are essential parts of evaluation and represent a significant feature thereof.\*
- (3) It appraises the "whole child" or the whole educational situation. It is not only the gathering and analysis of data but also the placing of value and reaching conclusions.
- (4) It is concerned with important functional learning outcomes, many of them less tangible and less easily measured.

Evaluation is partly the result of changing views in education. Emphasis has shifted from merely pouring information to meeting the needs of individual children as well. Often neither the child nor his parents nor his teachers

---

\* There is considerable disagreement on this point. Russell (16) found that "the statement....'self-evaluation is an important part of all evaluation' would seem to need both further investigation and extreme caution in its application." He found that there is a lack of scientific study in this area. Periodic evaluations of elementary school children bear little relation, he says, to those of peers and students. Self-evaluation (2) may provide better motivation and be more "democratic" but there is little evidence that it results in better learning.



know just why he didn't do well and how he could be helped to do better. Teachers now stress the importance of knowing what each child is like and why he is that way (3) This cannot be accomplished by the traditional testing methods only and requires a variety of instruments. For maximum effectiveness it must be an integral part of the major tasks of a school.(4)

b) Purpose: The ultimate purpose of evaluation is to improve the educational progress of the school and to make it more efficient (7) This is done basically by appraising the progress of pupils toward realizing the educational program. It answers such questions as what progress are we making toward meeting the objectives of the educational program, are the methods effective, are we really changing behavior, are the objectives achievable, are they worthwhile? The immediate purposes are

- (1) To validate the hypothesis upon which the institution operates.
- (2) To make periodic checks on the effectiveness of the institution, the maintenance of standards, the effectiveness of teaching or other elements of the program and to indicate necessary improvements.
- (3) To provide information basic to the effective guidance of individual students in their studies or problems of adjustment.
- (4) To effect the placement of students in homogeneous groupings and ranking and to predict success in a subject.
- (5) To motivate learning.
- (6) To develop more effective curricula and educative experience.
- (7) To provide an adequate and objective basis for measuring, recording and reporting progress in facilitating the desired learning and improving instruction.



(8) To provide psychological security to the staff. Uncertainty breeds insecurity.

(9) To provide a basis for public relations (4), acquainting the community with the progress of the school and answering critics.

c) Tests, Measurement and Evaluation

A test is "an examination or quiz; any kind of device or procedure for measuring ability, achievement, interest, etc." (10) Educational measurement is "a broad term for the general study and practice of testing, scaling and appraisal of aspects of the educational process for which measures are available." (10) Evaluation is an overall term which includes both testing and measurement as well as other means of appraisal of every aspect of that which is evaluated. Testing is, therefore, included in measurement and measurement in evaluation.

There was always testing. "Tests which purport to measure the efficiency of the work done in schools have always been used. Systems of examinations have determined not only the advancement of the individual pupil in our educational systems, but the results of these tests have at times been used as a measure of the teacher's efficiency and to determine the amount of public support to be given to a particular school or school system." (9)

Toward the end of the nineteenth century we saw the beginning of a movement for the objective study of education. (11) This was marked by Joseph Mayer Rice's study on "The Futility of the Spelling Grind" among 33,000 children over a period of sixteen months, released in 1897. Another important date in this development was the issuance in 1905 by Binet and Simon of their first composite intelligence test, followed by Thorndike's



achievement test. It culminated in Thorndike's famous dictum in 1914, "If anything exists it exists in some amount, and if it exists in some amount it can be measured." Educational measurement, implying objective, scientific means of testing on a large scale was now an accepted fact. The assumption grew that "educational processes could be standardized and their products "measured" by wholly objective and impersonal means. Once this was accomplished, guesswork would be abolished, teaching would become certain and sure and teachers scientific."(12) The emphasis in measurement was on accuracy and reliability . The multiplication of measurement which tend to be quantitative descriptions of objective conditions became the order of the day. It served to stimulate a tremendous amount of research.

The continuous study of education has brought about the realization that measurement, while a great step forward is, by its very nature, most inadequate. The success in the measurement of many quantitative aspects of education has also led to the search for procedures to measure qualitative elements and to extend it to a wider range of educational objectives. For measurement is tied up with objective testing where the emphasis is placed almost exclusively upon pupil achievement in subject matter areas. Objective tests, whose widespread use began in 1920 (13), emphasize single aspects of subject matter achievement or specific skills and abilities. It does not, however, necessarily imply evaluation which emphasizes broad personality

\*

---

The term "measurement" did not as yet appear in Paul Monroe's A Cyclopedia of Education, published in 1913.

\*\*

Measurement means the counting of something. Implicit is the assumption that through measurement we have more appropriate and more dependable information.



changes, stresses pupil growth and involves not only the collection and analysis of data but also the placing of some value on it or the reaching of a conclusion regarding its worth. (7) Going beyond measurement, evaluation assumes a purpose or an idea of what is "good" or "desirable" from the standpoint of the individual or society, or both. (1) Here attention is directed toward the broader aspects of the educational program.

Evaluation is, in a sense, a return to subjectivity, a realization that no scientific, infallible tests are possible in the study of human values. Education seeks to change the behavior patterns of human beings, it is more a process of growth than of acquiring skills and information, and objective testing is inadequate or unsatisfactory in evaluating progress toward this objective. Additional means, though less scientific, may prove more efficient and efficacious. This view is believed to have originated as part of the Progressive Education revolt against the traditional curriculum.(1) As increasing emphasis was given in education to goals other than knowledge and skills, the search began for ways, in addition to the standardized achievement tests, of appraising the educational program, to include such attributes as attitudes, interests, ideals, ways of thinking, habits and personal and social adaptability.

d) Problems of Evaluation

There is a considerable amount of confusion in this area. There is an increasing tendency to use evaluation, testing, and measurement as synonymous terms (7). In part the aims of evaluation have been incorporated in standardized tests. In part they receive limited use for two reasons: a) Schools generally haven't agreed on objectives, or they are vague, and purposeful education is impossible unless we know what we are trying to accomplish, unless there is a



relationship between ends and the means by which they are to be acquired(14). Often where there are goals they are limited to growth in skills and knowledge. b) Obtaining the evidence and summarizing the results is complicated(4). While objective tests are readily available and can frequently be used by the teacher without special effort- except perhaps selecting from the ever-growing variety, the procedures of evaluation are complicated and often require adjustment to individual circumstances and results may be indefinite and may require interpretation. Lack of funds and personnel are often a major deterrent, as is expert know-how. The tendency is, therefore, to retreat to appraising the structure of the school or school system and the progress in acquiring information. We may sell the real values of education short because of the apparent infallibility of procedures which emphasize only part of the process. This confining of programs to measurement of learning outcomes accessible only to objective appraisal by means of tests is, however, no longer acceptable (15). Nevertheless, the term 'evaluation' has had different significance in the work of different writers. This is so because different people have been concerned with different kinds of educational phenomena and with different aspects of the evaluation process.

- (1) There is an unconscious assumption that little can be accomplished in acquiring or changing interests; a corollary to this is a feeling that in most cases one interest is as good as another.
- (2) There is common association of evaluation with penalties either to students or to staff where serious deficiencies are revealed. Actually, should there be grave problems it is often the program which requires change.
- (3) There is a suspicion that those who set up interests have adult criteria in mind. We must be careful to consider the interests of the particular level.



e) Steps in Evaluation

The following are the steps in a careful process of evaluation:

- (1) Formulation of Major Objectives- This can be accomplished by (a) The curriculum analysis method, breaking down the general purposes of the curriculum into individual objectives, (b) The conference method, having a committee indicate the objectives to be realized and (c) The questionnaire and interview method, where a questionnaire aiming to discover objectives is used with students, parents, alumni or similar groups, followed up by interviews to check the validity of the questionnaire.(5)
- (2) Definition and clarification of skills and behavior to be realized in each objective.
- (3) Selection and construction of instruments and determination of their appropriateness for the various objectives. In the case of objective testing of learning information and skills the problem in general education is one of selection from the large variety of tests; in evaluating qualitative objectives of pupil growth it will be almost impossible to find ready-made measures.
- (4) Application of various formal and informal tests and techniques to obtain the information required for evaluation and interpretation of results.
- (5) Interpretation of results
- (6) Summarization to give a picture of the general trends with respect to every item studied.



## 2. Standards of Evaluation

A test or evaluative technique is generally judged on the basis of the following commonly-accepted standards: validity, reliability, objectivity, norms and practicability.

### a) Validity

Relevance or validity is the degree to which a test measures that which it claims to measure. This is most important. If a test lacks validity everything else is meaningless. (35, p. 200) To be valid there must be a close relationship between the test and what it is to measure. This is not an absolute characteristic- perfection is almost impossible- but rather a relative one, depending on the purpose. The same technique may be used for different purposes with different degrees of validity. Obtaining validity is beset with many pitfalls, and is at best qualitative.

There are four types of validity and the type should always be indicated (34) in discussing the subject:

(1) (Curricular or) Content Validity- The degree to which the test measures the material it is supposed to cover. The content of the test is here reviewed to determine if it includes a representative sample of the immediate objectives of the subject of study. This is of central importance in achievement tests. Of course, we must first determine what comprises satisfactory content of instruction- this is largely a rational analysis. Unless a test measures fairly the objectives of the instruction it may actually be harmful. The value of curricular validity is progressively decreased the greater the area of knowledge covered. The validity is specific to a particular curriculum or a particular task. This type of validity may be established by rating of teachers or experts or in



comparison with school marks. One national achievement test producer states they check validity against many textbooks and courses of study. It can be useful in analyzing achievement, plan remedial work and determine supervisory needs of teacher.

(2) Concurrent (Status) Validity- The degree to which the test measures more on less immediate behavior. It is the relation of a test to a current criterion, such as comparing a spelling test with actual spelling found in the children's compositions. A personality test diagnosing personal adjustment or an opinion questionnaire, interview, sociometric analysis or check lists would require concurrent validity. Other tests could help establish this form of validity. The difference between concurrent and predictive validity is solely a matter of time.

(3) Predictive Validity - The degree to which predictions can be accurately made from the test. This type of validity is necessary in predicting later success in aptitudes or vocations and is essential in tests of intelligence. The most common means of checking predictive validity is by correlating test scores with a later criterion measure of job success with the same or similar groups. Giving the same test or questionnaire two years in succession with allowance for growth factor may give a test predictive validity. We may want to predict if a student will succeed in language or science and give the test in the fall and compare it with teachers' marks at the end of the year. A basic procedure is to give a test to a group entering some job or training program and later to obtain for each one some criterion of success on the job or program and then compute the correlation between the score and the criterion



measure, with the criteria of success determined by a group of experts.

(4) Construct Validity- The degree to which we may make logical inferences from indirect evidence when direct evidence is unavailable. For instance, do seventh graders tend to score lower and ninth graders higher than eighth graders? Do pupils who have dictionaries in their homes have higher scores than those, otherwise apparently equal, who don't? Evidence in this type of validity is more varied and more theoretical (36). It is the degree to which a test is related to other characteristics or traits assumed to exist. We may correlate the results of tests of various mental or personality characteristics. We may establish construct validity by considering different kinds of incomplete but complementary evidence. Tests of these functions are valid insofar as they behave in a manner in which we may reasonably expect them to behave. We may formulate theories on measuring traits and they are valid to the degree to which they are borne out.

In seeking validity we must avoid face-validity- something which looks valid, particularly to laymen, but reflects inadequate or superficial analysis. It may reproduce facts but not broader measurements or talk about something but have limited relevance in actual behavior.

The first requirement for validity is that the test writer know what he wants to test. At times this is quite clear without spelling it out. If not the aims must be clearly indicated in advance. We then seek validity from

- a) the plan of the test: The method of determining its scope and contents and the provisions for review and
- b) The test itself: does it meet the aims of the plan.



Greater validity may also be obtained when the test a) has a large enough random sampling in each area to be tested, b) the items are well-constructed and free from ambiguity, c) The test is clear enough to read, d) it has enough choices (four is usually best, three acceptable) e) it has examples where necessary (usually two are preferred) f) it gives advance information on scoring where it would make a difference and g) there is a careful review of the test and the manual.

Thus far there seems little problem, except where there is need for stating and clarifying the aims. The basic problem, however, is that validity cannot be completely determined by the tester himself. Three outside elements are essential to validate a test and it is these, which we call external validation, which make the difference between a standardized test and others, between a good test and one difficult to determine. These elements are

a) statistical studies which cannot be determined through other means: (1) item analysis and (2) correlation of results with outside criterion and how substantial are the correlations, and for what types of groups

b) professional- through review by competent educators

c) outside criterion

(1) student marks

(2) opinion of teachers based upon observation over extended period of time

(3) agreement with grade advancement: the percentage of correctness should increase from grade to grade

(4) agreement between single question and entire test-this last is the most objective and perhaps the most reliable (41, page 60).



Obtaining satisfactory criterion measures is perhaps the most difficult aspect of validity. There is often no objective record of performance, and success in the work may be influenced by a variety of factors. Criterion measures are, therefore, only partial, and other elements considered. A World War II testing handbook states, "Generally speaking, the validity of the test is best determined by using common sense in discovering that the test measures component abilities which exist both in the test situation and on the job." (34) Unlike reliability, testing validity does not require giving the test.

#### b) Reliability

Reliability is the degree to which an instrument is consistent in repeated measurements, the variation which results when the same or an equal test is repeated to the same or similar people. A certain amount of variation is almost always inevitable. Will the results in another testing be almost the same? This is called the reliability coefficient.

The major reasons for differences in reliability are (a) the degree of a person's ability to take tests or to understand instructions (b) problems due to ambiguities and misunderstandings and the degree of difficulty of a test. If a test is very difficult there is much guessing and low reliability- and the reverse. c) health, motivation or environmental factors (d) degree of practice of the skills involved in the test (e) chance factors in administration or appraisal (33) (f) the number of items in the test- the more items, the more reliable is the test- up to a point of diminishing returns.

There are basically three types of reliability:

(1) Coefficient of Internal Consistency: The consistency in performance on



different parts of a test taken at one sitting. The test is divided for scoring purposes only into two halves, usually the odd numbers into one and the even numbers into another, and the scores of the two halves are correlated. It is the most widely used procedure because of convenience but it has limitations. (a) It does not reflect variations in individual from day to day (b) It becomes meaningless when it is highly speeded. Some prefer this form with adjustment by the Spearman-Brown formula. Any split-half test is somewhat arbitrary. Another procedure for correlating internal consistency is by item analysis. An estimate of test reliability may also be made from a single test administration by the use of the Kuder-Richardson formula\* if all items are given identical weights. It assumes that the items within one form of a test have as much in common with one another as do the items in that one form with the corresponding items in a parallel or equivalent form. If sound it would be similar to a coefficient of internal consistency.

(2) Coefficient of Equivalence- consistency of performance on two parallel tests, as equivalent as possible, with same number, types and difficulty of questions. The problem is to make them truly equivalent and the best guarantee for its accomplishment is the advance preparation of detailed test specifications. It is best to allow an interval of several days between the administration of the parallel tests. This method is preferred.

(3) Coefficient of Stability- The consistency in performance when the identical test is repeated after a distinct interval. Memory and a feeling

---

\* See Appendix D



that the person has done it before would affect reliability.

Reliability is but a means to an end and can add no importance to something trivial. One may overvalue reliability at the expense of the real significance of what is being measured. This may happen because reliability is often the most basic statistical data since validity is at times determined by rational rather than empirical methods.

Published tests try to get a reliability of at least .90. Good teacher-made tests should have a reliability of .60 to .80 (37). One national achievement test producer states they insist on minimum reliability of .80. If the reliability is unsatisfactory it may be improved by increasing the number of items proportionately, as follows: If you have a reliability of .40 and want to increase it to .60

$$\frac{.60 \times (1 - .40)}{.40 \times (1 - .60)} = \frac{.60 \times .60}{.40 \times .40} = \frac{.3600}{.1600} = 2\frac{1}{4} \text{ times}$$

Correlations range from 1.00 which is perfect, down to .00 where there is no relationship, then to -1.00 where the relationship is completely negative. It is meaningless to correlate scores of different groups, as for instance boys and girls.

A report on reliability data should cover at least the following (a) The operations upon which the estimate is based, b) A detailed description of the elements which might affect the reliability coefficient, c) The statistical characteristics of the group, especially the number of cases, the mean and the standard deviation, d) the adequacy of the sampling employed- it should be wide and representative, e) type of reliability coefficient computed, <sup>and</sup> f) reliability for single age and grade groups. Validated parallel forms give greater assurance of reliability when the same test is administered year after year.



c) Objectivity

Objectivity is the degree to which the personal bias and opinion are eliminated in reaching a decision. In a highly objective instrument identical or similar scores will be obtained when the same data is reviewed by different equally competent scorers at different times. Objectivity is generally preferred, and group tests in intelligence, achievement and aptitude are usually provided with scoring keys of high objectivity. Objectivity is related to reliability.

Some individual intelligence tests such as the Binet and the Bellevue-Wechsler have moderate objectivity, permitting the examiner some individual judgment in interpreting responses. Limited objectivity also prevails in performance tests, check lists and rating scales.

In diagnostic and clinical work flexible objectivity is especially desirable. The major evaluation methods in this group are projective techniques involving analysis of handwriting, drawing or the like; a running account of behavior by direct observation, anecdotal records and the open-end questionnaire.

Scoring keys help in decreasing errors and thus make a test more objective (they are also much more practical). Other features in validity and practicability also help make tests more objective.

A test should be as objective as possible without sacrificing validity. To help gain objectivity (a) a test should have clear, large type, well spaced (b) students should obtain advance information on scoring so as to decrease guessing, and other phases of the test, (c) examples should be given, (d) scoring should be objective - types of questions which can be marked only one way, and simple, and there should be a scoring key to decrease bias (also much more practicable.)



d) Norms

How do students compare with others in the same school, community or a larger area? Norms help answer these questions. A score is meaningless unless it is compared with something. A mark in a class may be in comparison with a standard the teacher sets up. This is, of course, the lowest type of comparison, as no two teachers mark the same way or base marks on the same criteria and no two classes are the same. Thus a pupil doing work of similar difficulty may receive different marks each year, especially if he changes teachers. One teacher may give an A and another a C for identical achievement. To overcome this problem and to help establish more or less valid criteria for comparison, norms have been established. The norms should be adequately prepared and listed in the manual.

A norm is a score obtained by giving a test to a representative group of people and comparing later individuals and group tests with it. Standard tests usually have norms based upon a fairly large cross section of pupils in widely scattered areas. One who takes a standard test can thus compare the results with the norm group. Norms should refer to clearly defined groups and be based on a representative cross section of the group. A small number of students selected at random in a large number of schools may yield more reliable norms than many more students in few schools (37) and a selected smaller number of schools with all types of ability than a larger number chosen at random (Acorn). State and local norms are often more useful than national norms. In local situations we may have modal norms, eliminating retarded and accelerated pupils. Meaningful comparison require appropriate norms.

The major types of norms are age norms, grade norms, percentile norms and



standard score norms. Age and grade norms are derived by an average of wide representative sampling and individuals and groups are compared with them. A child who had an age norm of 10-5 is one who has the ability of an average child 10 years and 5 months old, although his actual age may be only 8. What is called an I.Q. is obtained by dividing one's mental age by his chronological age and multiplying it by 100. In the above illustration his I.Q. is  $\frac{10-5}{8} \times 100 = 131$ . Where a third grader has the norm of a fifth grader it does not necessarily mean that he knows the work of that grade. It may mean that he gains the status by obtaining a near perfect score in third grade work while the norm which consists of an average is only 50%. The meaning and significance of age and grade norms are subject to question because of unequal growth and study in different periods and numerous differences between group and group but they probably represent the best method available.

Percentile norms compare a person with his own age and grade group. It is a point on the scale below which a given percent lies. Instead of indicating that the 8 year old has the ability of an average 10 year old we say that he is in 95th percentile of his group, meaning that all but 5% of the students have grades below his. This is particularly useful in setting up local norms. There should be at least 100 scores for a percentile norm.

Standard score norms have been developed because percentile units are unequal. There are usually many percentile units in the middle and few at the top and bottom. A change of 5% in one is different from the same percentage of change in the other. The standard score is a unit that is equal and has the same meaning throughout the scale. The standard score is the number of standard deviations above or below the mean. It expresses the position of a score both



in relation to the mean and to the variability. If the standard deviation is 5 and the mean (expressed as 0) is 50, then a score of 65 is expressed as +3.0 and a score of 30 is -4.0. Percentile ranks are generally preferable to age and grade equivalents for direct comparison and standard scores to percentile ranks. The standard deviation is considered the most accurate measure of variability. The problem is that it's assumed to be rather complicated.

A norm and a standard are not necessarily identical. A norm is an average of a representative group; a standard is an acceptable score, a desirable model or goal, and may be considerably above or even below the general norm. If a class has outstanding students only or is retarded we cannot be content by comparing their results on a test with a norm created by testing students of all types locally or nationally. Teachers often confuse the two and believe a class is doing well if it is up to the norm, regardless of other factors such as general intelligence and curricular deviations. "While norms may serve as a guide, the evident purpose of estimating pupil growth is to compare what he has done with what he's doing now, his actual competencies with his own potential and latent abilities." (69)



e) Practicability

In writing or purchasing a test one should consider different phases of practicability:

1. Economy of Time- a) Saving time through use of adequate scoring key and equal weights for different parts of test, b) Use of fewer categories in preparing the test and not too long for one sitting and for administration- not to complicate marking. However, decreasing time of the test through the use of fewer items is often false economy because reliability depends on the length of the test.
2. Economy of Cost- a) Low initial cost of test, b) Possibility of reusing test booklet by having separate answer sheets- it also saves much time in scoring. It has been estimated that students spend up to twenty-three hours taking tests and \$20 on fees each year. \*
3. Availability of Staff- Degree to which teachers can be used for administration, scoring and interpretation. Standard tests should not be used if their administration cannot be mastered by average teachers (5). It's preferable to have scoring done by one member of the staff. Equal weights and answer sheets help considerably in providing practicability.
4. Ability to Interpret Test.- There must be a manual which should include a) Purpose and method of development, b) Clear directions for administration and scoring, c) Information on type and method of determining validity, reliability, and norms and coefficients and scores, d) How test results are to be used, e) How it compares with other tests, f) Groups for which valid, g) Age, grade, size and nature of group.<sup>(1)</sup>
5. Comparability- the existence of parallel forms of the test which are helpful for administration before and after a particular period of study or to confirm a questionable score.

---

\* New York Times, 2/19/61



## SUMMARY OF STANDARDS FOR TEST EVALUATION

### I General Information

- a) Name of test
- b) Date issued
- c) Author(s)
- d) Grade(s) Intended
- e) Text(s) or Other Content Reference

### II Requirements for Test Evaluation-

We may state that there are 21 requirements, of unequal weights. Some are interrelated and may be placed in more than one category. For our purpose we may list them as follows:

#### a) Validity

##### 1. Internal

- (a) 1. Set the aims and make sure the test meets them
- (b) 2. Have large enough random sampling in each area
- (c) 3. Well-constructed items
- (d) 4. Clear enough to read (also objectivity)
- (e) 5. Guessing factor adequately handled
- (f) 6. Have examples (also objectivity)

##### 2. External

- (a) 7. Item analysis
- (b) 8. Professional review
- (c) 9. Correlation with outside criterion

#### b) Reliability

- (a) 10. Enough items for reliability
- (b) 11. Well-established parallel forms (also practicable)
- (c) 12. Tested for reliability

#### c) 13. Norms

#### d) Objectivity

- (a) 14. Questions which can be scored objectively
- (b) 15. Scoring keys to decrease bias (also practicable)
- (c) 16. Advance information on scoring (also validity)



e) Practicability

- (a) 17. Adequate manual
- (b) 18. Equal weights
- (c) 19. Reasonable length
- (d) 20. Few categories
- (e) 21. Answer sheets

Where degree of a particular standard is to be judged, four categories will be listed: A- very well B- satisfactory C- poor D- not at all



### 3. Instruments of Evaluation

A variety of instruments have been developed for appraising all phases of child growth, both quantitative and qualitative. The acquisition of knowledge as well as the development of personality. A variety of uses may be found for each technique, administrative, instructional, guidance and research. These uses overlap because categories are not mutually exclusive and the same data may serve multiple purposes.

The evaluation techniques may be classified into the following somewhat arbitrary categories: (a) Objective Tests, constituting one of the major techniques of evaluation, (b) Essay and oral examinations, frequently used by classroom teachers as informal methods for assessing and diagnosing day by day growth, (c) Observation techniques, including anecdotal records, (d) Questionnaires and inventories, (e) Check Lists and Rating Scales, (f) Interviews, (g) Projective Techniques, (h) Cumulative Records and (i) Case Studies.

The first six groups and case studies are among the major evaluative devices (24) Almost all also use cumulative records. Standard achievement tests are administered regularly in most school systems. In the typical system mental tests are given 4 or 5 times during the child's school career.

In school and community surveys the same instruments are usually used, omitting non-objective tests and adding experiments (27). In evaluating features of school systems the following are used: comparative procedures (with other units in system, other systems, average and outstanding practice); application of test standards, score cards and rating scales; measurement against research results, judgement of survey staff, expert opinion and check against trends.



a) Objective Tests

An objective test is " a standardized instrument designed to measure objectively one or more aspects of a total personality, by means of samples of performance or behavior" (29). Psychological, mental, educational or new-type tests have as their basic feature short answers and objectivity in scoring. They are, therefore, also called objective or short-answer tests. Objective testing means that the tests are marked objectively, not that they are prepared objectively and certainly not that they are based on objective standards of achievement for "there is virtually no objective and accurate basis for determining standards of achievement in school work... It is not possible to state, for example, in terms of experimental evidence that after a pupil of a certain age, ability and background has experienced a certain amount and type of learning, he should have acquired certain specified learnings." (31, pp. 762-3)

The basic advantages of an objective test are

- (a) It can review much ground because it can include many questions
- (b) It has a high scoring reliability
- (c) It is easy to mark
- (d) It enables identification of pupil weaknesses
- (e) It enables review of test for greater retention
- (f) It pinpoints what is actually new in a new lesson

The basic weaknesses of an objective test are

- (a) Difficulty of construction- it requires more time and effort and some skill
- (b) There is a degree of guessing in most types of questions



(c) The major weakness is that it doesn't readily lend itself to development of thoughts (In teacher-made tests an essay question or two is, therefore, often desirable.) Most objective tests measure only retention of information and not judgement or attitudes.

Tests may be divided into various classifications.

a) Intelligence Tests- Tests that tend to emphasize abstract intelligence and predict general scholastic achievement.- Language aptitude tests are superior to general intelligence tests in predicting achievement in foreign languages (85) While these were found helpful in predicting general school achievement, they cannot be depended upon to predict achievement in specific subjects, especially in the study of foreign languages.(86) Intelligence tests scores have shown up disappointingly both as regards validity and reliability. A summary of a dozen studies shows a correlation between intelligence tests and grades of from .12 to .69, with a median of .48, the median being only 13% above that afforded by chance. Intelligence test results were proven less valid predictors than a combination of the pupils' grade school record and teachers' marks, which averaged above .60 (87). One study found a close relationship to intelligence in extremely weak achievement, .65, but a negligible correlation in high achievement and found such discrepancies between mental ability and achievement test that she questioned the validity of predicting language success on basis of mental tests (88). Spoerl states that there seems to be rather general agreement that intelligence is not the only factor involved in learning a foreign language and that average marks correlate more closely.(89) Von Wittich found that the highest correlation



with foreign language, .73 is with total grade average, that English grades are second highest, with the I.Q. on the bottom with a correlation of .48.(90) Tallent found a correlation with English of .558 and with the I.Q. of only .211; and notes that while correlations in educational testing almost invariably fall below .60, a coefficient of .50 may be said to indicate probable relationship. Most tests fluctuate between  $\pm .25$  and  $\pm .50$  (91). "Binet and Simon, and Terman studied their tests as predictors of scholastic achievement. Their findings are similar to those revealed by an extremely large number of subsequent studies, using various tests and various criteria of scholastic achievement; namely the correlation commonly falls in the range of four to five tenths..." (2, p. 717)

b) Aptitude and Prognosis Tests- Predict success in a particular subject or vocation. These contain questions not directly dependent on school learning. Some question if these tests can improve prediction over a combination of achievement and intelligence tests (35) One may have a high scholastic aptitude and do poorly for other reasons such as attitudes, effectiveness of instruction, interests, intensity of motivation, home environment, studiousness, outside activities and personal-social adaptation. The validity of the prognostic tests is usually low. There is also the question as to whether there is a special aptitude for foreign languages not measured by general intelligence tests. No one has been able to delineate this purported ability (82). Assuming there is, the question remains how it is determined. Two studies include English vocabulary, stating that there is a high correlation between the two (90,91) and one of them includes arithmetic. Others feel that there is a relation between one foreign language and another. Still others say that to predict in one



language one must include questions in that language alone. On the college level one study found that tests in language aptitude are superior to intelligence tests in predicting achievement in foreign languages but not sufficiently high to assure accuracy in case of individual students (86). Other significant information such as teachers' marks should therefore be taken into consideration. Actually, very little work was done in language prediction.

c) Achievement- Tests of learning, concerned with evaluation of past progress. These are specifically planned to measure what's learned in school. The difference between achievement and aptitude tests is one of degree. The achievement test score is most often relied upon as the criterion for placement of pupils in homogeneous classes. It might prove interesting to know just how many very bright children have been in "slow classes" on this basis (93)

d) Interest Inventories- Questionnaires to determine vocational interests and aptitudes.

e) Personality measures- Tests of character, temperament, adjustment and attitudes. Though usually regarded as tests they are more in the classification of questionnaires or inventories, as they lack the element of right or wrong. Fortunately, a high proportion have appropriate attitudes and these tests therefore present important evidence. Personality is measurable if we accept it as a more or less definite entity, as the sum total of the individual's behavior in social situations. We have the same problem here as in case of intelligence, where tests have definitely been useful (18). Tests of achievement in certain fields provide much evidence on personality. The validity of almost any tests on personality has not been established. Its value as testing is greatly curtailed.



because it depends on the veracity of responses. The reliability of most personality tests is about as high as many of the widely used intelligence and achievement tests but few are as high as the most reliable of them.

## 2. Differences As To Purpose

a) Survey Tests- "Provide a general, overall appraisal of status in some area of knowledge or skill." (35)

b) Diagnostic Tests- "Provide a detailed picture of strengths and weaknesses in an area." These are generally confined to low achievement groups and the tests have many less difficult items. They require reliability evidence for each subject and must therefore, be much longer than survey tests. Construct validity is widely useful here. These tests are not widely used and few are published except in reading and arithmetic and these have meager evidence of reliability and norms.

## 3. Differences As To Area Covered

a) Standard Tests: Intended for wide scale use and emphasize general objectives and common components appearing in different texts. They test courses that are fairly uniform nationally. They are valuable in measuring growth and development over long periods of time and for providing comparative data based on relatively uniform standards of measurement such as age or grade. They are usually prepared by experts, by groups rather than by individuals, have high reliability and satisfactory norms and item analysis. They are not directly related to teaching in any single school or grade.

Standard tests are subdivided into (1) Single Tests- different forms of one test for one subject and one age group and (2) Batteries- a group of tests in different subjects and different grade levels. The typical battery has from



four to ten tests. The chief advantage of batteries are unity in norms and in plan and convenience. They are generally felt to be advantageous despite rigidity and other problems including (a) local emphasis may differ (b) population characteristics vary (c) causes teacher to stress those skills which the tests emphasize over others which he may consider more important. Batteries help to (a) identify pupils for more detailed study (b) plan program of class and obtain good general picture and (c) create homogeneous groups. They should not be the only means of judging.

b) Local Tests: City or state-wide tests for specific objectives of instruction to fit a particular course of studies. These are often superior to national tests because the population is more restricted, less varied and more known and the course of studies may be fairly uniform. They also enable better tests and greater ease of administration than teacher-made tests and better motivation than national tests. They also make for comparability of records from one school to another and one year to another. They provide a greater identification of teacher with test than do standard tests.

c) Teacher-Made Tests: Help to motivate pupils, to determine achievement and to encourage proper study habits. They should aim to have validity, reliability, item analysis and other qualities found in standard tests. Building a file of successful test items is helpful in this direction. In practice, however, "The usual school grades offer an inadequate and sometimes an invalid and unreliable measure of the student's abilities and potentialities". (31, p. 762-3) The characteristics of a good test are the same for a formal printed test as for an informal test made by the teacher (46)



#### 4. Differences As To Those Tested

- a) Individual Tests: Generally refer to intelligence tests or personality measurements administered to each child individually. These are used for guidance purposes, as they are costly and time-consuming. It seems that they do have a higher reliability than group tests.
- b) Group Tests: Standard group tests are generally repeated every 3-4 years and have a correlation from grade three through high school of perhaps .50 to .60. These tests may be considerably shorter than individual tests.

#### Planning the Test

##### 1. The Preliminary Form

- a) Determine clearly the objectives of the test- they must be related to the content and be achievable. The test should be necessary.
- b) Outline the material and decide on the abilities to be tested, the subdivisions and the approximate number of items.
- c) Write the items, placing those of similar content together.
- d) Assemble the test. If time permits put it aside for awhile before checking it.
- e) Prepare practice exercises and set time limits.

##### 2. Pretesting

In preparing a test one should usually make about twice as many items as he will ultimately use. This should be reviewed by a committee of experts (41) and the preliminary form of the test is then given to a large experimental group and the answers to each item analyzed. This is item analysis which is the "basic operation that all published tests have to undergo and the basic reason for whatever superiority they possess." (37) This consists of three parts:



- a) Success- Maximum reliability is obtained if every item is answered correctly by 60% to 70% of those tested. A question should normally be excluded if more than 90% or less than 30% answer it correctly. An item should be neither too easy nor too difficult. Investigators have proven that precision in measurement is greatest when all items of a test are about equally difficult.
- b) Discrimination- A good test item is usually one which has at least 20% more correct responses in the upper half (or quarter) of the class than in the lower half (or quarter). Thus, in a class of 40, if 6 of the lower half answer the question correctly, then there should be at least 10 correct answers in the upper half (20% of 20). We must be careful toward the end of the test since low-scoring students may not have reached this point. The analysis should be dropped after a fifth of the students have dropped out. One writer suggests comparing the best third with the poorest third- Those items be considered most valid which show the highest percentage of successful response in the highest third as compared with the response with lowest third. (46)
- c) Effectiveness of Distractors- Options which no one chooses should not be included. If high ranking students tend to choose distractors or the reverse, the item should be reexamined. The final form is then drawn up and given to thousands of pupils of a representative population for norms. Content validity is evaluated by specialists and teachers, reliability and other validity studies are made and the manual is prepared.

Objective tests include questions of recognition and of recall and the items may be divided into four major categories: (1) multiple choice, (2) matching, (3) true and false and (4) completion. In every type of item one should make sure



that (a) instructions and questions are clear (b) questions are concise and within same group, of uniform length (c) it doesn't include minute details which students shouldn't be expected to know (d) except where memorization is desired statements are not copied directly from text. Some specific suggestions for the various categories follow:

(1) Multiple Choice- The multiple-choice question is more objective in its scoring and also enables avoiding absolutes by indicating a best answer. The alternatives should represent common errors, related material or misconceptions, and the difficulty of each item depends on the closeness of the distractors. It is the most popular form of item in standard tests. Four choices seem most common and on a secondary or higher level five may be preferable.

(2) Matching- is actually a form of multiple-choice. Here <sup>it</sup>/is desirable that one side have more statements than the other and it is helpful to have each set of items of a more or less homogeneous nature; short matching sections help avoid confusion. Sections of five questions and seven choices are advised.

(3) True-False items suffer from the fact that statements are often neither true nor false. This form of item is least valid because it offers only two possibilities. To discourage guessing in true-false questions and to a lesser degree in the others, test makers often deduct the number of wrong from the number right.

The formula is 
$$\text{Right} - \frac{\text{Wrong}}{\text{number of options minus one}}$$

Thus if the section of the test has 30 items with five options in each and the person answers 20 right and 8 wrong, then it's  $20 - \frac{8}{5-1}$  or 18. The special manner of grading should be indicated in advance.



(4) Completion or Short-Answer Items are the least objective because there are frequently several possible answers. One must therefore be careful to see that the blanks call for simple answers. These are not widely used in standard tests (33).

In preparing a test it is, finally, preferable to have few sections with many questions in each.

Robert W. Travers in "How To Make Achievement Tests" gives eight steps in planning evaluation instruments:

1. State goals in general terms
2. Define goals in specific terms
3. Assign weights to goals
4. Outline course content
5. Prepare blueprint which organizes activities
6. Use blueprint to prepare specific questions
7. Add to test plan other pertinent item such as scoring systems etc.
8. Select measuring techniques (type of test)

(80)



b) Non-Objective Tests

Non-objective tests consist of essay and oral examinations:

1. The Essay Examination is an important and widely-used means of evaluating achievement. While it is often used because it is comparatively easy to construct and many teachers are wary of using other techniques, it has at the same time some definite educational objectives: (a) It seems better equipped to test the functioning of one's mental powers, comprehension of sense material, relationships and interpretations, (b) It may judge one's ability to organize material effectively (c) It may show one's originality of thinking and fluency of style (d) It may provide greater insight into one's learning and personality.

The major problems of essay examinations are (a) consistent scoring unreliability. There's disagreement among different scorers and even in the same scorer at different times. Different experts have scored the same questions all the way from excellent to worthless (41 Appendix A 1-8 and pp. 37-41). There's the question of what to measure and the division of credit and such extraneous factors as one's ability for self-expression, comparison with others and the tendency to grade the student as well as the paper. Because careful scoring requires a lot of time, it's more affected by the temperament of the teacher. (b) Narrow range of material covered since only a comparatively few questions can be covered. (c) Disproportionate emphasis on the correct interpretation of each question since there are so few of them. (d) No diagnostic value, since it is practically impossible to trace a child's real shortcomings from his answer to a complicated problem. (e) Children who don't know the answer are often inclined to bluff their way through by generalizations- leads to dishonest habits.



Much of the criticism of the essay test can be obviated by careful construction and grading: (a) Each question should be planned to measure a definite objective of instruction for which no valid or reliable short-answer test is available (b) The question should be given in detail to require a specific and restricted answer (c) Choices should be avoided because they make comparisons impossible (d) A standard answer should be formulated with specific credit for each point expected. Extra credit may be given for clear organization. (e) In order to avoid a "halo" effect evaluate all responses to one question before going on to the next one and anonymity may also be helpful (5)

A good deal of perfectly valid evidence can be obtained from essay examinations if they are carefully organized and graded (2) This, however, is usually more difficult than in objective tests and would eliminate the very reason many use this form of testing.

2. Oral Examinations- This presents the weaknesses of the essay examinations and has the further problems of (a) how to prepare questions of comparable weight to different students (b) asking even fewer questions of each one (c) consuming too much time (d) devoting very little time to each student and (e) requiring spontaneous often subjective scoring. It is a good means of studying the thought process of a student, diagnosing his difficulties and helping him think through a problem. It's basically a teaching rather than testing technique. It is, of course, necessary for the younger child and helpful to draw out the timid one.

The oral test can be used in interviewing teachers. The interviewer should see that (a) questions are clear (b) they are specific, requiring a specific answer (c) they are related to the subject and test what is actually necessary for the teacher. Ten to fifteen questions should suffice.



c) Observation

Observation is the oldest and commonest instrument of scientific research (22) and is used in studies of practically all aspects of behavior. It may be divided into (a) structural or controlled observation- obtaining data on previously defined activities and (b) unstructured or informal observation- observation made by the teacher in the classroom without previously set pattern. It involves a systematic recording in objective terms of actual behavior as it occurs in a natural situation in a manner that will yield quantitative individual scores.

Direct observation gains in reliability and validity when (a) it is restricted to limited selected aspects (b) equal time, periods and similar conditions are given to each (c) systematic and objective records are kept. The procedure, however, fails if objectivity and reliability are gained by sacrificing the substance with which the study deals.

One method of observation (6) is for a trained person to occupy an inconspicuous place in the room and record each activity which falls in any of the defined categories. This is scored either quantitatively, dividing the number of entries by the number of days and pupils or, qualitatively, rating entries by several judges and then dividing the score by the number of raters. Wrightstone (23) gives the following techniques in studying pupil participation in planning the curriculum: (a) a number of classes are visited where pupil planning is in progress and notes made of participation (b) These are then organized into categories (c) The tentative code is applied to several classrooms to observe its functioning (d) the code is then revised.



A basic advantage of observation is that with certain precautions, it permits measurement without disturbing the normal activities of those observed. The data can be treated by conventional statistical techniques. A problem in direct observation is that it generally requires a trained investigator, takes much time and is costly. There is also the problem of fitting the observer into the picture, eliminating subjectivity and bias, determining meaningful procedure and interpreting it. To aid in observation, charts, check lists and other instruments are used.



d) Anecdotal Records

Anecdotal records are a form of observation involving the recording of anecdotes concerning specific, limited aspects of pupil behavior which seems significant. It's a description of the conduct and personality of a limited number of children by means of frequent, brief recording of incidents made by the teacher. It is <sup>an</sup> informal method of teacher evaluation of pupil growth and adjustments to help them understand and guide children. They are most valuable in areas of social and emotional adjustment. They provide a basis for rating of children and offer validating evidence for evaluation instruments. The incidents are important in their cumulative value by themselves or together with information gathered by other means.

To be effective (a) There must be enough anecdotes to enable evaluation, (b) they should be observed in typical situations, (c) they must be accurate and objective, brief, factual and clear, so it can be understood later, (d) they should contain statement of situation in which anecdote occurred (e) entries should be dated and recorded on the same date. The teacher should concentrate on limited aspects for which information cannot otherwise be obtained. There are several important "don'ts." Don't give (a) personal reactions (b) primarily dramatic incidents (c) interpretation before adequate facts are obtained (d) don't worry if significant items are omitted- this is not intended to be an inclusive catalogue of incidents. Observations should generally be made in situations in which child has freedom to display a variety of reactions. Most teachers may need a year to gain experience in writing and interpreting anecdotal records.



Trailer (18) indicates need for (a) enlisting the cooperation of the teachers, making them understand that their effectiveness as teachers depends on the knowledge of the behavior of the individual child (b) setting rules for observation, including a reasonable number of anecdotes he's to record each week, preferably after a week or two of experimentation. It is advisable to select a specific aspect of behavior such as cooperation or punctuality, otherwise it's too haphazard. (c) preparing simple records for writing down the anecdotes. In the course of a semester the teacher should have several anecdotes for each child. There should also be a record for summarizing the anecdotes. The record may simply be a mimeographed card listing the name of the student, the date, the place, the incident, and the name of the observer. It may also have place for comments as to conditions under which the incident took place and the interpretation.

The anecdotes take on added significance if several teachers make records of the child. This may include the homeroom teacher, guidance counsellor and principal. The teacher should review his notes every month or so to check on trends and decide when he has sufficient information to make conclusions. At the end of the semester he should be ready for tentative interpretation.

Since they are highly time consuming anecdotal records are usually supplementary in a comprehensive evaluation program, used to help make more meaningful other more systematic but less colorful type of evidence.



e) The Questionnaire

The questionnaire is used to obtain facts to elicit problems and to ascertain opinions and attitudes. One form of the questionnaire is the inventory, a more or less arbitrary subdivision in which one is asked to give short or "yes" or "no" answers to a rather complete set of questions. There are two types of questionnaires, one in which the answer is checked or selected and the other an "open end" form, allowing free responses.

The reliability of the questionnaire is more subject to question than of the interview because it lacks the personal factor. This is corrected at least partly by the comparatively large numbers reached by the questionnaire. Its validity is partly dependent upon the types of questions and the manner of their administration. Where possible it should be tried out in preliminary form with a small group to clear up ambiguities. Greater validity may be obtained by (a) limited number of questionnaires among competent people (b) definition of terms where judgement is involved (c) reliance on averages and other measures of group opinion.

There seems to be little difference between signed and unsigned questionnaires unless the questions are of a highly personal nature (19). Even in such cases students may convey the information if they have confidence in the teacher. The signed questionnaire has the advantage of allowing follow-up and is a must when the purpose is diagnostic. Objective information about physical conditions of the home appear to be reported accurately (5).

The questionnaire may serve a major purpose in the educational process by helping determine conditions regarding the objectives of the curriculum in order to seek means for implementation, correction and change. When used at



the beginning and the end of the year the questionnaire may reveal changes which have taken place in individuals or in the group. A good questionnaire must have clear purposes which become the criteria of what to include in it. Its usefulness in evaluating students depends on (a) the degree to which it samples activities covered by the curriculum and (b) the sincerity of the response; this can be helped by making the students realize that it's not a test and that the information may be useful.

One form of the inventory is the personality inventory; a collection of questions and statements designed to yield data on the individual's social and emotional adjustment. While the individual doesn't always know aspects of his personality or is unwilling to reveal them, this is nevertheless a valuable form of locating poorly adjusted pupils. One inventory tries to assess the strength of various needs or motives in the life of the individual. Interests show a good deal of stability, at least in adolescents and adults and such inventories for the college age can compare favorably in stability with intelligence tests.

The questionnaire and the inventory are personal reports when the individual describes or rates himself and his reactions or behavior.

A good questionnaire must be carefully constructed and grouped: (a) The questions should be clearly stated, (b) they should be few and relevant to the purpose (c) they should be relevant to the individuals questioned, (d) they should be easy to answer, (e) the questionnaire should be well designed and numbered, (f) decide in advance how it's to be tabulated. This will be reflected in the questions, (g) it's preferable to have questions prepared



by more than one person. In all cases the questionnaire should be reviewed by at least one other person.

A major problem in using the questionnaire is obtaining a large percentage of replies. The following may help: (a) There should be an introductory statement on purpose to elicit cooperation (b) there should be specific directions for answering questions (c) an effort should be made to select topics on which the recipient wants to know the answers (d) it should be sent to people who know you and respect you and a copy of the results should be promised (e) there should be a vigorous follow-up.

The questionnaire doesn't require a large staff or costs and can save much valuable time. It may be effective where people may record on paper what they wouldn't tell orally. It can be used in the school, among numerous uses, in comparing values of children of different grades and backgrounds. It can be especially useful in the following instances:

- a) Obtaining information in a temporary group which may not be available later
- b) When people can't be reached personally
- c) When there are too many to be interviewed
- d) When data are either non-existent or not conveniently available
- e) When information is desired in a group, giving each privacy in replying and eliminating one's answer being influenced by that of the other.
- f) As a jumping-off place for an interview



f) Check Lists and Rating Scales

Check lists and rating scales are similar instruments of evaluation. They are used by evaluators to judge the presence of certain qualities, previously determined, in an individual or institution. One may check the most applicable answer from two or more choices or the presence or absence of certain qualities, or the degree of their presence (always, most of the time, frequently, rarely or never). Sometimes there is a forced choice, when one must choose between two alternatives which are approximately equal. The check list judges others while the questionnaire gives one's own facts or opinions about things.

A check list becomes a rating scale when the evaluator makes a judgment of the degree to which a certain quality or fact is present: it quantifies judgments. It becomes a rating scale by the addition of a scale of values, which may be comparative or expressed numerically or otherwise.

Check lists and rating scales are usually used for evaluating a total school program or various aspects thereof. It may be used to rate books, a course of studies, methods which are successful or enjoyable, goals to be attained, characteristics desired of teachers, students or groups, desirable activities of children. There are also diagnostic check lists to check difficulty of pupils in certain areas. Pupils may be rated by counsellors, teachers or parents. In observing behavior one should be rated independently by at least two teachers in a position to observe the child.

One should not turn to rating scales except as a last resort (28). The reliability of this method is improved as the number of judges and their expertness is increased. The correlation between the sets of ratings is then



computed. Greater objectivity is gained in preparing the check list in consultation with others; it also increases the likelihood of wider application. In many cases it is helpful to have students have a hand in preparing the check list; they are more likely to favor it and it is less likely to have errors. The self-drawn check list must by comparison be subjective, narrow and prejudicial.

Unlike the questionnaire the rating scale can be used to evaluate one's own reactions and can also be used by children who can't read a questionnaire. One of the difficulties is a "halo effect"- rating assigned to one trait influences rating on others. One way to get around this is to mark one trait on all papers, then go to the second trait, etc.

Thurstone has developed the following technique in constructing attitude scales:

1. Collect statements on a given social institution from many sources.
2. The statements are sorted by a large number of judges into several groups with respect to degree of favorableness.
3. About 20 statements are then selected ranging from extremely favorable to extremely unfavorable and a scale of values is assigned to them.

Common faults in check lists are (a) too many items (b) overlapping items (c) too many or too few steps- seven steps will yield optimal reliability in most instances (5). These may result from too few raters- at least three are advisable; personal bias; and failure to use extremes if scale-raters often use middle scales only.

Check lists, rating scales and questionnaires, are also used for sociometric purposes- presenting relations among members of a given group and time. This includes choices of associations, people with whom one wishes to do certain



activities, and qualities one finds in different people. The results are usually presented on a sociogram-pointing arrows in the direction of those chosen by different individuals. It is used to identify and help children in social adjustment.

In obtaining sound ratings, there is the question of willingness and of ability to do so. The teacher may want to give the student a break and if promotion is involved he's likely to say "excellent" unless he wants to dispose of him. A teacher who may have 100 students may not be able to answer questions on initiative or originality. The principal and the teacher may not view the student from the same vantage point. And how good is "good" or how fair if "fair"? The ranking of students may help overcome a tendency to want to give a good rating to everyone. Symonds concluded (35) that correlation between ratings of two independent raters is .55.

Check lists and rating scales have a degree of objectivity in that they remind different observers what to look for and to look for the same thing and the scale makes the judgment of different observers more nearly comparable.

A check list requires only recognition, not recall, and is therefore less difficult. On the other hand it is often only a sample.



g) The Interview

The interview is a direct method of obtaining information through person-to-person questioning. It is used in studying individuals and in obtaining the views of a limited number of people in the study of attitudes and to obtain possible meanings and causes behind objective factors (17). It may produce deeper insight of complex problems than is possible by means of other procedures.

If time permits it's the best procedure for gathering information (18) and in a small school it may replace the questionnaire. Some feel it is "an extremely valuable means of conducting evaluation in the school." (5) Others, however, question the advisability of the time and expense required to train and employ interviewers when questionnaires and similar techniques are only slightly less reliable (25).

There are three types of interviews:

- a) Diagnostic- obtaining facts and opinions on a person
- b) Survey- obtaining facts or opinions on a problem, as in a public opinion poll.
- c) Treatment Interview- seeking to help individual adjust to a situation

The advantages of the interview are

- a) It is flexible, allows clarification and follow-up questions not previously considered but resulting from the interview's development at this point.
- b) It tells not only what the person said but how he said it which may be of greater importance
- c) One may learn important unrelated facts from an interview such as emotional reactions
- d) Some confidential information can be gained only by this technique
- e) Some people can't or won't answer questionnaires or other techniques



The interview presents many problems

- a) The interviewer requires special training and personal qualities. He must be informal and cordial, must establish rapport with his subject and must use varied approaches.
- (b) It is difficult to ascertain validity and reliability. Correlations between estimate of student ability obtained by interview and actual marks range all the way from - .66 to .73 (17) The evidence for validity is spotty and contradictory (35)
- (c) It is a costly, slow and hence a limited process
- (d) Because of its free and partly spontaneous nature it is more likely to include much wasted time and unnecessary and irrelevant material and ignore some obvious areas.

Another problem of the interview is that taking notes may have an adverse effect on the interviewee. Some write summaries, trying to recall questions and answers after the interview. One student found, however, that only 10 to 35% of the interview is covered in this manner. Some take brief notes and summarize the main points later. There is an increased use of recordings although this procedure may place the interviewee under special tension. One partial solution is a structural interview, maintaining a written guide while still retaining freedom and flexibility.



#### h) Projective Techniques

Projective techniques are methods by which an individual projects himself into a situation outside of himself and thus unconsciously reveals himself. They make possible a qualitative and comprehensive study of the individual and hold promise of deeper insights but are valuable only in hands of specialists. Some forms present less difficulty in interpretation but they yield correspondingly less revealing analysis.

Perhaps the best known of these techniques are The Rorschach Tests of inkblots which can also be given to groups through slides. Another is Murray's "Thematic Appreciation Tests"- The interpretation of a series of ambiguous pictures on which one is asked to tell or write a story. Both can be used by specialists only and are intended to reveal facets of one's personality. One may reveal himself also in drawings and paintings and in his use of play materials or in the completion of sentences, stories or plots. It is an area of tremendous variety but is extremely difficult of interpretation.

The autobiography may be used as a projective technique. In such case a student should receive a list of things to be included and must feel sure their confidence won't be violated. Of course, one may exaggerate or minimize matters and a home visit and an interview may improve validity.

In the interview, questionnaire and inventory we see what individual says about himself, in rating scales what others say about him, in objective tests and observational techniques we see what the person actually does, how he behaves in the real world of things and people and in projective techniques how he behaves in the world of fantasy and make believe (35).



i) Cumulative Records

The cumulative record is a compact, convenient and comprehensive summary of significant information on individual pupils during the years of their stay at the school. It should list interests, activities, accomplishments, experiences of note associations, social adjustment and personality ratings, health, mental and emotional factors, aptitudes, talents and attitudes. It should have extensive space for test scores. The most complete cumulative records also include personal information (even photo), enrollment and attendance, home and community influences (even economic status, church affiliation, language spoken, home duties and influences) and supplementary material (anecdotal records, case studies, psychologists' notes, teachers' comments)(24)

Although rarely mentioned prior to 1925, the cumulative record has become a dynamic force in the individualization of the work of American schools (20). The record gives one a feeling of familiarity with the student. It enables the teacher to get acquainted quickly with new students, helps to plan a program that fits the child, to identify problems, to gain data for guidance, for conferences with parents, for case studies. It plays a unique role in the long range evaluation of pupil growth. Its predictive value is increased when data on an item is gathered annually over a number of years. It has a naturally high predictive value because a number of related items will usually provide better prediction than a single item. Nevertheless, since the number of possible entries is limitless, each entry should be subject to the question: what will it contribute toward diagnosing the child?

The cumulative record may be a card, booklet or folder, with preference for the latter. The booklet such as used in the Denver schools, has place for



entries and ratings by child, parent and teacher. It is kept cooperatively. The folder, as used in New York City and in California, has the advantage of providing space for essential supplementary materials not previously planned for. Diederich (28) suggests the addition of a "profile index." Here the various objectives of the school are listed and the teacher records next to each objective the percentage of growth achieved.



j) The Case Study

The case study is a detailed analysis of an individual involving the integration and use of comprehensive and significant data. It is a combination of multiple instruments, the synthesis of all pertinent facts on an individual. It is the "most scientific method now known" for the prediction and control of behavior (21). It is an "indispensible" aid in bringing about better adjustment now that schools began emphasizing mental hygiene and guidance (18). It should be used for children with serious difficulty in adjustment or scholarship. It can help gain greater insight of an individual and help him develop his potentialities to an optimum degree.

The case study may include an introduction giving the reasons for the study, of intelligence and achievement tests, learning defects and personal problems and information gathered in the cumulative record. It may include interviews with or written statements from the child himself, friends, parents, teachers and supervisors. It may require additional tests, observation.

Some case studies end with diagnosis, others report treatment. If it is for a special purpose, such as difficulty in reading, then only those items which are relevant to the problem are gathered. If it is for general improvement in adjustment then every item which may be important is included. Unskillfully used it becomes a meaningless chronology of guesswork and misinterpretation. Desirable characteristics are completeness of data, continuity, confidential recording and scientific synthesis.

The case study often requires a psychologist. It depends on the area



and the difficulty of the case. Studies in reading difficulties, for instance, have often been successfully made by teachers. Treatment of difficulties out of one's experience should be referred. If the study lends itself to measurement comparable tests should be given at the beginning and at the end. There should also be follow-ups after the completion of the study.

### The all-day school

As the degree that Jewish education is provided in all-day schools, and it is combined with regular studies under one roof and under Jewish auspices, all instruments applicable (in general) education are also relevant here. The Jewish all-day school is limited in this direction only by (a) the admission and retention policies, usually not providing special facilities for children with very low I.Q.'s or presenting severe emotional adjustment problems, and (b) its limited student bodies and number of such schools in a community, with the possible exception of New York City, where the existence of religious exemptions or religious exemptions and psychological and religious exemptions is often the subject of public school policy. On the other hand its responsibility is more encompassing than that of the public school because it combines both religious and secular studies and the child's religious upbringing and religious education. It is generally considered that the majority of the religious staff of the all-day school. This would, on one hand, establish a project and social adjustment in the child's life, and on the other hand, establish a religious and social adjustment. It would, on the other hand, intensify the need for adjustment and social adjustment and religious adjustment and religious adjustment.



### III. APPLICABLE TECHNIQUES IN EVALUATING JEWISH EDUCATION

We have indicated thus far the crucial need for evaluation in Jewish education in general, what constitutes evaluation and what instruments have been developed in evaluating general education. We must now investigate the specific techniques applicable and essential in evaluating Jewish education, survey the existing instruments and their use, indicate their shortcomings and make proposals for their improvement.

#### 1) The all-day school

To the degree that Jewish education is provided in all-day schools, where it is combined with secular studies under one roof and under Jewish auspices, all instruments applicable in general education are also relevant here. The Jewish all-day school is limited in this direction only by (a) its admission and retention policies, usually not providing special facilities for children with very low I.Q.'s or presenting severe emotional adjustment problems, and (b) its limited student bodies and number of such schools in a community, with the possible exception of New York City, making the services of guidance counsellors and psychologists difficult to acquire. (The problem can often be solved by obtaining part-time services of public school personnel or of a Jewish case work agency (67,68). On the other hand its responsibility is more encompassing than that of the public school because it combines both the secular and religious studies of the child and specific moral training is generally considered more the province of the religious than of the secular school. This would, on one hand, eliminate observation, projective techniques and case studies, usually used in special individual cases and requiring trained personnel. It would, on the other hand, intensify the use of interview and questionnaires and render valuable all other instruments of evaluation.



2) The supplementary Jewish School

In considering evaluative instruments for the supplementary Jewish school, either afternoon or Sunday, we must decide on (a) their need and (b) their practicability. Generally speaking, it must be recognized that the needs of the child in the supplementary school are similar to those in the public school. There are, in fact, added problems due to the voluntary nature of this type of institution. More than 30 years ago, Julius B. Maller said, "It is true that conditions in the Jewish religious school differ from those in the secular school. But the differences are largely in aim and philosophy. There is no reason why the modern Jewish school should refrain from adopting the best practices, tools and devices of secular education. In fact, the greater complexity of the religious educational aims makes more cogent the need of appropriate instruments to determine whether or not these aims are being realized. (41). This was true then and is as true today. At the same time it is unnecessary to duplicate that which is done adequately in the public school, when that information is available, as may be in intelligence testing. The maintenance of anecdotal records, used largely to study social and emotional adjustment, may be unnecessary because it may duplicate and also impractical because of the limited time of the teacher with the pupils, greatly curtailing this evaluative effort. Observation and anecdotal records could be practiced on a limited basis in areas considered the particular province of the Jewish school. Furthermore, since the child's stay in this school is limited, the institution's responsibility in such areas is also limited. We cannot expect social and emotional adjustment to be the responsibility of a school where the child spends less than 10 to 20% of his total school time during any one year and much less in the overall school span. It must, at best, be limited.



Those instruments not applicable in the all-day school would also not apply here. All other evaluative techniques may be used in the supplementary Jewish school.

3) The aims of the Jewish School

In order to decide which specific instruments are essential in the Jewish school, we must know what the Jewish school aims to achieve. We can then determine how we can evaluate the accomplishment of these achievements. As in the case of Mitzvot, the aims of Jewish education can be subdivided into numerous categories; they can also be condensed into three or even into one overall purpose. "The goal of Jewish education," says one noted educator, "is...the living of a traditionally Jewish life." (61) Another says the same in different words

התורה היא חיינו (62)  
החיים הם התורה

These statements, both by representative orthodox Jews, are all-inclusive generalizations which would probably be approved by all groups in Jewry, eliminating the words "traditional" and חיים הם תורה and subject to interpretation and elaboration. Another statement, by a Reform rabbi, which would be generally acceptable only if supplementary to the above, is, "Religious education must lead not only to information and knowledge but also to character growth and personality development." (63a) Another rabbi says, "Of all the goals for which we aspire in the Yeshiva there is one which overrides them all - instilling moral values in our children. The Yeshivah which may have excelled in imparting a profound knowledge of Torah without moving the souls of its students toward moral behavior is a dismal failure in its historic mission." (73) It will be agreed that this is true of all religious education.

Obviously, the purpose of learning has always been not merely to acquire knowledge per se but also and more so to develop and enhance the human personality,



in a sense to change the individual. The Bible expresses this in the aim "to do righteousness and justice;" the Midrash, commenting on the verse, "...and may be gracious unto three," says, "May He grant you knowledge so that you be gracious and compassionate to one another;" and the Talmud says,

(Kedushin 40b)

היה רצון, ויהיה רצון  
שהיה רצון, ויהיה רצון

Indeed all three groups in Judaism have more elaborate and detailed renditions of the aims of Jewish education. The aims of Jewish religious education as enumerated by the Reform wing in Judaism (63b) are seven in number. The Conservative group \* lists eight objectives (64) and an authoritative Orthodox statement of aims (65) can also be broken down to eight. These and the seven elements enumerated by Dr. Dushkin (66) may be taken as the composite of the aims of the American Jewish school \*\* excluding in some measure the very small number of secularist schools and those of an anti-Israel and anti-Hebrew persuasion. As Dushkin and Engelman state: "There is striking similarity in the official statements of fundamental or guiding principles and aims in Jewish education made by all the commissions representing the main "denominational" organizations in American Jewry. There seems to be very little indeed in the statements of one group that the others object to."(42,p.32)

This is not to imply that the details in the various aims are the same and that the American Jewish school is monolithic. The Reform statement of

---

\* For a detailed study of the objectives of the congregational school see "Issues in Jewish Education" by Louis Katzoff, Bloch 1949

\*\* See Appendix A for comparative aims



practices differs radically from the Conservative and the Orthodox, and the Orthodox interpretation of Torah differs vastly from that of the other groups. There are vast differences in the degree of study. Both the Orthodox and the Conservative espouse the Day School but, judging by results, the Orthodox are much more devoted to it. The Reform state that Hebrew is an "indispensible element...and must play an important role in our course of studies." but the fact is that when a new Reform school is established it will most likely be a one-day school with Hebrew occupying much less time and being taught to much less numbers, and on a voluntary basis, than schools established by the other groups.\* The same may still be said of the teaching of Jewish history and of post Bar Mitzvah and girl training in the average Orthodox school. The general aims, however, are the same and they may be stated as follows:

a) Learning

- (1) Torah, Hebrew, history and other texts, taught for their place in Judaism, the transmission of our religious and cultural heritage and the instilling of Jewish loyalties
- (2) Knowledge to enable the practice of Jewish ritual: participation in Synagogue service and home ceremonies- ability to read and acquaintance with the Siddur, putting on Tefillin, conducting a Seder, knowing the laws of kashrut, Sabbath and holidays.

---

\* The Jewish Education Committee of New York stresses these differences in its school subdivision. "It was possible to get schools to operate on a common curriculum pattern and develop a program of testing from such a curriculum most effectively by organizing schools along ideological lines."  
(98) One wonders, however, if the curriculum and testing features were the factors in this development.



b) Practice

The observance of mitzvot, in accordance with the interpretation of the particular group, including

- (1) Torah- study as a religious principle and as a lifelong process
- (2) Avodah- prayer and other Jewish observances  
participation in Jewish life  
moral and ethical living
- (3) Gemilut Hassadim- giving Zedakah  
working for good causes

c) Values and Attitudes

- (1) Love of God and faith in Judaism  
Reverance for Jewish traditions  
Appreciation of Jewish heritage  
Feeling of kinship to Jewish people, past and present (Self-  
identification with things Jewish)  
A feeling of joy in being a Jew and in practicing Judaism  
Appreciation of the place of Hebrew in Jewish life  
Kinship toward the land of Israel  
Positive attitude toward Jewish study  
Positive attitude toward prayer  
Favorable attitudes toward Jewishness
- (2) Spiritual and ethical sensitivity  
Love of one's fellow man  
Interrelation of Judaism and democracy



(3) Development of one's character

Jewish standards in meeting personal problems

Training to be a happy, healthy individual and a well-adjusted

Jew- having a sense of security as a Jew

4. LEARNING

In the area of knowledge it is essential to have information on

a) The child's ability to learn (1) generally through intelligence tests and (2) specifically through aptitude and prognosis tests, especially in language learning; and

b) The degree of the acquisition and retention of the material studied: achievement tests.

(a) Intelligence Testing: Are Intelligence tests necessary in Jewish education? These tests are essential at all times in order to compare potential with actual achievement. This is even more essential in Jewish education because here there is more often a definite gap between ability and the desire to learn, between ability and motivation. It is also helpful in classification in ability grouping, although Nardi (85) found a correlation of only .31 between I.Q. and success in Hebrew School and Mrs. Ettinger found a high correlation in grades 2-4 (.462 to .712) but not in grades 1 and 5 (.247 and .109) (82) Such tests would assist the teacher and the supervisor to know when the child is not doing his optimum and in studying problems of individual pupils. It aids in programming for acceleration or for remedial instruction. In the supplementary Jewish school, if the information is not obtainable from the Board of Education, such tests should be given regularly during the first year in the Jewish school and every



three-four years thereafter. The question of availability of intelligence tests is not relevant in Jewish education. There are ample tests meeting the desired standards available in general education and they can be used, when necessary, in Jewish education. No special intelligence tests are necessary in Jewish education.

(b) Aptitude Tests: Almost all afternoon and day schools and a large percentage of Sunday schools, together comprising the vast majority of children studying in Jewish schools, devote some time to the study of a second language, usually Hebrew but at times Yiddish. It is generally agreed that a noticeable percentage of these children have little aptitude for a second language and waste their time studying Hebrew, and yet we make little effort to separate these children and to prepare for them a program of studies commensurate with their abilities. This is partly due to a feeling that the study of Hebrew should be pursued by all, that it is indispensable in the Jewish School, to a fear that its elimination will further weaken and decrease the program of studies in the Jewish school and that the slow child is slow in everything and we may as well not teach him at all. In a large percentage of schools this is so because they are too small to allow a double program and with this limited time available it is often felt that separation within the class is unwise. However, there are many schools which can cater to the special needs of these children as well as to those of exceptional ability and don't. Where ability divisions are made they are usually done after the passage of a year or more, based solely on the evaluation of one teacher or on the basis of ability in general studies. The first causes the waste of a year of the child's limited time in the Jewish school and is often too subjective to be accurate. Judging solely on the basis of



success in one class or with one teacher is often inadequate because a child may do well with one teacher and poorly with another. The second is also at times inaccurate because of unequal abilities in different areas of study. Prognostic tests in language aptitude are essential to further aid at a proper classification. They are also useful where a teacher feels that a child is doing considerably below his ability.

(c) Achievement Tests: The evaluation of the acquisition and retention of Jewish knowledge, both learning per se and learning to practice Jewish traditions, is by far the easiest and most efficacious of the three areas, but by no means a simple process. In order to determine our success in this area we must first analyze our course of studies and delineate the major areas of learning.

(1) Divisions in the Jewish School: The Jewish school may be divided into three areas: primary-through the age of 7, where achievement testing is not applicable; elementary-ages 8 to 13, where the bulk of our children study prior to Bar or Bat Mitzvah and which generally constitutes the elementary Hebrew school in almost all Conservative and Orthodox Schools; and the high school, which includes a comparatively small percentage of the Jewish-school population.

The Jewish school may be divided into four groups: the day school, the Hebrew school, the Yiddish school and the Sunday school. The Hebrew school includes the Congregational school, which now represents the bulk of our educational institutions, and the communal Talmud Torah; the two are generally identical in program, although the Talmud Torah is usually more intensive or, to quote a prominent educator, "The Congregational school is a smaller replica



of the Talmud Torah." (70) The day school is an extension and intensification of the Talmud Torah, with little difference in the subjects of study in the lower grades and added subjects in the higher grades. Even the Sunday and the Hebrew school, judging by the official courses of study, have a similar program, except in degree and in the use of Hebrew, whether they be Orthodox, Conservative or Reform. Except for the Yiddish school, "while they differ in time allotment and emphasis (they) exhibit a striking similarity of pattern." (70)

(2) Subjects of Study- General - Thus, as in the case of aims, there is much greater unanimity in the program of the elementary Jewish school than is generally realized. The major courses of instruction in the Conservative school are Hebrew, prayers, laws and customs, Humash and history. The Orthodox school has less Hebrew and more Humash and the study of history is uneven, with the other two subjects about the same. The Reform school may be one day or several days a week and may or may not have Hebrew in the one-day school. The other subjects in the elementary grades are history and customs and ceremonies. While the most recent Reform curriculum outline does not list the study of the Bible in the elementary grades, a survey of Reform schools (72) indicates that the subject is taught in a large number of its schools on this age level as well. The numerous courses prepared by national organizations, Bureaus of Jewish education and individual schools differ largely in the period in the student's career when a specific subject is introduced, in the amount of time devoted to the subject and in the textbooks used, rather than in actual differences in material of study. One course prescribes the teaching of Humash to begin in the second year, another in



the third or even later, some cover more ground and others less, different schools devote different amounts of time to the subject, but virtually all Orthodox and Conservative schools teach the Humash. The same is true of the other subjects. The same is true of the Reform school, except for a much more limited Hebrew program and the Bible being largely taught in English rather than in Hebrew as is the practice in all Hebrew schools. Bureaus frequently prepare identical courses of study for schools of different religious groups. One educator writes, "Two separate curricula, one for the Conservative and one for the Orthodox, were originally contemplated. However, the rabbis and the principals of the two types of schools found that with minor changes and adaptations the same curriculum could be used by both." (71) Of course, no two programs are identical; there are differences in details, emphasis and interpretations. Both may teach the Bible, but the Orthodox and the Reform will interpret the Story of Creation, or Miracles or Revelation in totally different ways. Very great are the differences in stressing the Sabbath, holidays and other observances.

(3) Subjects of Study on Elementary Level- The major differences between the Reform and other Jewish schools on the elementary level are the teaching of the Bible in English rather than in Hebrew and the complete absence of Siddur instruction in Reform schools while it is taught without exception in all others.

A survey of 14 Courses of Study of all groupings and including most of the well-known programs for children until the Bar Mitzvah age shows 13 listing Hebrew, 12 Siddur, 12 Torah in Hebrew and in English, and all 14



history and laws and customs. \* Less frequently taught are music-9, personalities-9, prophets-8, Current Events-8, Ethics-5, Rashi-4, Mishna-2, Yiddish-2 and Jewish community-2. Personalities and Current Events are, of course, part of history. Thus if (1) we include the early prophets as well as the Pentateuch in the part of the Bible taught in the elementary Jewish school, we list (2) Hebrew, (3) Laws and Customs, (4) Siddur, (5) History and (6) add music we will include all the subjects listed in the vast majority of courses of study.

---

\* This is corroborated by the Survey of Jewish Education in New York City, 1951-52 (77, p. 24)



TABLE 1: SUBJECTS AND GRADES LISTED IN COURSES OF STUDY FOR ELEMENTARY GRADES

	Hebrew	LAWS & CUSTOMS	HISTORY	Siddur	Torah	MUSIC	PERSON.	PROPHETS	CURRENT EVENTS	RASHI	ETHICS	MISHNA	Yiddish	Comm. Life
United Synagogue	1-3	1-3	3-6	1-6	4-6		1,2		3-6					
Model Program for T.T. Jung-Kamminetzky	1-6	1-6	1-6	1-6	2-6	1-6		4-6	4-6	5,6	5	6		
Boston, 5-day schools	1-6	1-5	1-4	1-6	3-6	1-5	1	4-6		3-6				6
Boston, 3-day schools	1-3	1-6	3-6	1-6	3-6	1-6	1,2		1-6					
Shavitch Girls T.T. אגודת נשים חסידי	1-6	1-6	4-6	1-6	2-6	3,4		5,6			5,6		5,6	
אגודת נשים חסידי אגודת נשים חסידי אגודת נשים חסידי	1-6	1-6	3-6	1-6	2-6	1-4	1,2	5,6		5,6	5,6			
Essex County	1-3	1-5	2-5	1-5	4			5	2-5					
Yeshiva Torah Council Queens, Sampson A. Sweroff, 1960	1-5	1-5	2-5	1-5	2-5		1		1-5					
אגודת נשים חסידי אגודת נשים חסידי אגודת נשים חסידי	1-6	1,3-6	2-6	1-6	2-6	1	1	4-6		3-6	5	5,6	4-6	
Outline for the Jewish Rel. Sch. UAHC	2-7	3-6	5-7				3,4		4-7		7			
Outline for Cong. School Lang, 1951	1-3	1-6	4-6	1-6	4-6		1-6							
Course of Study Outline for Jew. Sun. Sch. - Chicago		1-2 4-6	3-7		7	1-7		7	6-7					
Outline for Heb. & Yiddish Sch. - L.A.	1-6	1-6	3-6	1-5	3,4	1-6	1,2	4-6						4,5
Course of Studies for 3-day Heb. Sch. - New Haven	1-5	1,5	2-4	1-5	3-5	1-5			3-5					



(4) Feasible Tests on the Elementary Level: Standard tests are feasible in all five major subjects of study, with tests in the Bible prepared both in Hebrew and in English, in order to meet the needs both of the Hebrew and the one-day school. Standard objective tests are also essential for all textbooks in all areas of study. Specifically, 1) Standard tests in the Hebrew Bible in the afternoon schools may be helpful for the first six Sidrot of Genesis, for the second six, and individually for each of the other books of Moses; also for each of the books of the early prophets, with separate tests for each of the books of Samuel and Kings. The abbreviated version, practically standard in these schools, should be used. In the English Bible separate tests should be prepared for each book and an overall test for the Pentateuch and one for the early prophets. 2) In Hebrew, standard tests are needed based on a composite of the major Hebrew texts used in our elementary schools. There should also be tests on each of these books and series of books, such as Haivri, Sippuri and the Scharfstein, Greenberg and Bridger texts. 3) In Laws and Customs, separate tests are feasible on the holidays and on the other laws and customs. These tests should all be in English and must be related to specific textbooks. 4) In Siddur we need one test for mechanical reading including speeded reading and another on the significance of basic prayers and the order of the services on different occasions. 5) In Jewish history, we need overall tests, and separate tests divided into three periods: from the beginning until the Babylonian Exile, from the Babylonian exile through the Spanish period and from the 16th century until today. Separate tests on Jews in America and on Israel are not applicable to this age level. Each of the volumes and the entire series of the Pessin, Gamoran,



Zeligs and Klapperman histories should have standard tests.

All tests, except those in Hebrew and in mechanical reading, should include both questions of fact as well as interpretations and reasons; they should be both quantitative and qualitative. Tests in the Bible and in Judaism should, where possible, emphasize moral aspects and interpretations which lend greater significance to the narrative.

We need not only tests which review the subject currently studied but also cumulative tests to measure retention of material studied in previous years, so that at the end of the third year we may also test the degree of retention of the material studied in the same subject during the first and second years; also we need instruments to test the degree of retention after the passage of a year, two, five or ten years after the person terminates his studies. In most instances the same cumulative test often given in Hebrew at graduation from elementary school or in history at the completion of the course, can also be given several years later. The problem usually is not one of test availability but of utilization. To be used for this purpose, however, the <sup>question?</sup> stricture of eliminating questions answered by more than 90% of the students must be disregarded. It may be those very answers will be remembered and this should not be considered as being of no consequence. In preparing or using a cumulative test one must also be sure that each phase of the subject which is to be reviewed is adequately represented.

The same problem exists in other areas. Tests may be used for diagnostic purposes, to analyze weaknesses in instruction. Tests which measure different aspects or periods of study may be used for such purpose, and some tests are. A Hebrew test which has an adequate section on grammar, on sentence structure



or on spelling, can be used to analyze success of instruction in that area; a history test which adequately covers different periods can be used to analyze the success of study in a specific period; and a test in Humash or Laws and Customs can be used to evaluate the teaching of moral and ethical principles, although a knowledge of them does not necessarily imply application or appreciation of these principles. In each of these cases it is not necessary to devise special tests but rather to be sure the problem to be analyzed is properly covered in the test and then to proceed with the analysis.

Tests are feasible as a means of studying the adequacy of our textbooks. Where agreement can be obtained by a consensus of knowledgeable educators on a word list in Hebrew; on selected passages, emphases and a series of interpretations in Humash; on a list of items to be included in a test on Jewish history, a test carefully geared in these directions can show the adequacy of the textbook as correlated with the criterion. This is, of course, predicated on a knowledge of and an agreement on the aims of teaching a specific subject and implies a comparison, through the test, of aims and results. In the case of teaching elementary Hebrew there is considerable agreement on the aims and selection of the Chomsky Word Lists and as such, when properly validated, the Chomsky test can be used to evaluate the adequacy of Hebrew textbooks, except where the book indicates different aims. Thus far no such criterion exists in other areas of study in the Jewish school. However, tests based on essential facts in all of Jewish history, such as that devised by the AAJE, may also be used to test the adequacy of specific textbooks in that subject. If average classes, using a specific text, consistently are considerably below average in



this test while doing satisfactorily on an adequate test based solely on that textbook, then it may be assumed that the textbook fails to meet the content validity established by the criterion. Again, it is not necessarily essential for this purpose to prepare special tests but rather to investigate the validity of a specific test for this purpose.

The above, of course, implies more or less identical conditions in all factors but the one tested. If radically different methods are used in one case, or the time element differs considerably while all other factors remain constant, then a test as just described would help determine if the special factor creates entirely different results.

(5) Subjects of Study on the Secondary Level:

a) If similarity is a distinguishing mark of the elementary Jewish school, the reverse is true on the secondary level. (a) There is a sharp division between the one-day school of all groupings and the Hebrew High School. While both teach Jewish history and laws and customs, the first adds a survey of the Bible in English, Ethics, Jewish problems, comparative religion, and a survey of literature, while the second adds Hebrew language and literature, the Hebrew Bible and its commentaries, and the Talmud. The Reform secondary Hebrew school adds the beginnings of the Hebrew Bible and prayer book, both included in the elementary departments of other Hebrew schools, and a course on Reform Judaism.

b) Within the Hebrew high school no two communities or schools seem to have a program which borders on the uniform. It is the reverse of the secular school which tends to be similar because of admission requirements to



colleges and College board examinations. A survey of seven courses of study reveals that they teach 9 different subjects and of these six are taught in three or more schools each.

Table 2: SUBJECTS TAUGHT IN SEVEN HEBREW HIGH SCHOOLS

	NEW YORK	CHIC. CONG. SCHOOLS	CHIC. COMM. SCHOOLS	BOSTON	HERZLIAH	BUFFALO	NEW HAVEN
HEBREW	1-3	1-4	1-4	1-4	1-4	1-5	1-5
MISHNAH	1-4	2	1-3	1	X	3	1-4
BIBLE	Early Prophets	1-3	1	1,2	X	1-4	1-3
	Latter Prophets	3,4	3	2-4	3	X	5
	Writings	4	3	1,2	4	5,6	4
HISTORY	2-4		1-4	2-4	X	2-4	
MISHNAH	4			1,2		5	
TALMUD			2-4	3,4	X		4-5
CODE OF LAWS	1,3,4		2,3	4	X	6	
PRAYER		1-4					
RASHI	2	2	2-4	1	X		1-4
MIDRASH			2,3				
MUSIC			1-3		X		

X -All grades

All teach Hebrew language and literature and Bible with commentaries; Talmud or Mishnah is taught in six schools (Talmud in 4 and Mishnah in 3), History and Laws in 5 schools. However, in Hebrew 29 different texts are used. While the basic word lists are a guide to the vocabulary used in elementary Hebrew school, no such aid is available on the secondary level. Some use Israeli



texts, others texts basically prepared for the American public high school and the Lador and Sifria Oneg readers, still others use ungraded more advanced texts. In the Bible, some use the complete text, others an abbreviation; some teach continuously, others use selections, and no two schools teach exactly the same material. Ten different texts are used in the eleven classes teaching the Talmud and three different texts in the four classes in the Mishnah. Only Laws and Customs and Jewish history present a degree of uniformity in the Hebrew high school and even in history one school begins from the Babylonian exile and another ends with the French Revolution.

A survey of textbooks used in ten Hebrew High Schools in different parts of the country (Chicago, '59) lists 28 in addition to seven books on Dikduk; 9 histories, eight complete and one on Israel; 9 different books of early prophets, including 6 different selections. The least variation is in Humashim, all abbreviated, and in Hebrew texts of Laws and Customs.

c) The one-day high school, which for the purposes of this study begins at the eighth grade in public school has a program of studies and series of texts no less uniform than the elementary Hebrew school. It is the only area where a uniform text is used almost throughout in the teaching of the Bible, in Comparative Religion, and in Modern Jewish Problems. (72) Few texts are used in the other areas except in Judaism and uniform standard tests can also be prepared in the latter without unusual difficulty. Single tests are needed in each subject and in the Bible there should be separate tests on the major books, one on the latter prophets and one on the Writings as separate units.

The Hebrew high schools can establish standard tests only in history and in Laws and Customs. In other areas of study each school must create its own tests for each text and each subject of study.



5. PRACTICES:

In the area of practice ordinary testing is, of course, not applicable. The easiest and perhaps most efficacious form would be the check list, to establish if the ways of life taught by the institution are practiced. In the Orthodox school, this may include if the student prays daily, eats Kosher and Passover food in the school cafeteria; in all schools it may include if he attends Sabbath and holiday services, is absent from public school on Jewish holidays, contributes to Jewish and general charitable causes and participates in the solicitation of funds for such causes, reads about and shows an interest in Jewish affairs, is active in Jewish club work, is respectful of his parents and teachers, is friendly and helpful toward fellow students of different religions and races, is honest in his dealings with his school and associates.

The problem is that the validity and reliability of such a check list is obviously not high, that the practice need not necessarily reflect training in the Jewish school and it depends to a large degree upon the home. Where the home is not observant the child can hardly be expected to be so. However, while the child of an observing family is more likely to follow this path than the child of a non-observing family, the religious school is a factor in determining his action. It is well-known that children are occasionally observant despite their parents; more common is non-observance by children of observant parents.

The necessary precautions should be taken (See section on Check Lists and Questionnaires) to assure a maximum of accuracy and the check list should be given at the beginning of the child's study in the religious school and after the passage of time, at the end of every two or three years in the religious



school. The check list should not be given by the children's teacher and, if possible, should not be taken by that class alone, to decrease the fear of identification. We must also devise different check lists for different religious orientations or make a list of uniform statements and add others for different groups.

The accuracy of the check list may be validated by (a) Teachers' opinion in the school on a group basis, (b) A group questionnaire to ascertain how many or what percentage participate in the activities listed above. Such a questionnaire is easy to devise and its validity should be great. The problem is that it would test not so much the personal commitment of the children as it would the successful organization of the school's program and the commitment and cooperation of the home. However, while individual items may be the result of outside influences the entire result would undoubtedly indicate a tendency and to some degree show the efficacy of the training. (c) An additional and more accurate check of the effectiveness of the training is to conduct a check list or questionnaire of the same or similar groups several years after the children had completed their religious training.

It would be helpful to have a check list or questionnaire of the home and its practices, the ethical and religious observance of the parents and the degree of imposing them upon the children. Such an instrument should be administered before the children enter the religious school, upon their completing the school program and several years later.

Ascertaining the degree of performance would help us (a) determine the degree of success in endeavoring to train for the practice of Judaism, (b) keep



us on the alert for our failings and our problems in this area of instruction and help us seek remedies and improvements.

6. VALUES AND ATTITUDES:

Evaluating the success of this area of instruction is similar to that of evaluating practices. Many of the values are interrelated with practices, result in practice or are a result of practice. A positive attitude toward prayer, study, traditions or Israel must somehow be related to practice. Others, however, need not necessarily be so, may be difficult to relate to practice, find little opportunity in that direction or be extremely difficult to evaluate. We may have to create hypothetical situations to test the individual's reactions. This may be true of checking on one's spiritual or ethical sensitivity, his love of his fellow man, his adjustment as a Jew or his feeling of kinship to his people.

The most satisfactory instrument in evaluating, attitudes and values is probably a multiple choice questionnaire asking the individual what he would do under certain circumstances, how he would feel if he had a choice between going to Europe or Israel, if he were asked to play the role of Jesus in a school play, if he had a choice of joining a Jewish or a Hy-Y club, if Negroes moved into his neighborhood, if he had a choice between cheating or possibly failing a test, if all the prayers were recited in English, if he were called a dirty Jew.

This form of questionnaire is (a) difficult to prepare, (b) must be reviewed by several knowledgeable individuals, (c) must have several related questions in order to consider varieties of the same situation and to spot discrepancies and irregularities, (d) must have many questions in order to increase validity, (e) must not be too obvious so the child won't answer what he



thinks you want him to answer. A committee of experts must determine the value scale of the possible answers.

In addition to such general questionnaire to be given at the beginning and end of the individuals years of study it would be helpful if shorter questionnaires are given or individual subjects such as honesty, brotherhood or attitude to Judaism at certain periods of the student's stay in school. Individual classes can be observed by an outsider in specific situations where one or the other areas particularly lends itself or the teacher can take anecdotal records in several areas over a period of time. Where attitudes of children are particularly objectionable it may be well to visit the home and interview the parents to try to learn the reasons for the attitude.

The results of a questionnaire on attitudes can be the determining factor in deciding on an additional course or unit on ethics or on Israel or another area of instruction, on the conscious introduction on frequent occasions of stories and discussions on specific values and attitudes which constitute aims of our educational program.



#### IV. SURVEY AND ANALYSIS OF EVALUATIVE INSTRUMENTS IN JEWISH EDUCATION

##### 1. Historical Review

a) Publications: The year 1910 is a starting point for a study of many areas of modern Jewish education in American for it was the year of the establishment of the first Bureau of Jewish Education. This is not so, however, in regard to testing. The first community survey of Jewish education in New York City, made in 1909, does not list testing as one of the essentials (43) and Dr. Samson Benderly, the architect and philosopher of the Bureau, makes no mention of the subject in his statement of aims and achievements issued in 1912. This is so despite the fact that he speaks of a standard curriculum for the large Talmud Torahs and says that in business stock is taken often and at regular intervals while "in Jewish education stock is taken every few hundred years, if at all" and "the Jewish people has not taken stock of its educational methods for many generations." (44)

The lack of discussion of or even reference to testing and evaluation in Jewish educational literature is widespread, almost universal. With but few exceptions there is virtually no reference to testing in books and periodicals on Jewish education until 1959. This includes Julius Greenstone's "Jewish Education in the United States"(1914), "Jewish Education in the United States, 1901-1950" by Leon L. Honor (1952), "Jewish Education in the United States in Mid Century" by Israel Chipkin (1951), "The Education of the Jewish Child" by Richard C. Hertz (1953), "Report on the Findings of the National Survey," United Synagogue (1950), "A Survey of 125 Religious Schools" by Emanuel Gamoran (1925), "Issues in Jewish Education" by Louis Katzoff (1949), and "The United Synagogue Commission on Jewish Education, 1945-60" (Hebrew) by Jay B. Stern.



Israel Chipkin's "Twenty-five Years of Jewish Education in the United States" (1937) has but one reference to the subject: "Definite attempts have been made to determine objectives to evaluate achievements through development of tests and measurements and to improve methodology." (p. 86) No further explanation or mention is made of the subject. "Central Communal Agencies for Jewish Education" by Abraham P. Gannes (1954) clearly indicates by the process of omission the paucity of testing in Bureaus. While it lists uniform examinations as a possible Bureau activity in the questionnaire used to obtain information for the volume, there is also no mention of the subject in the index and there are very few references to it throughout the volume, except a statement that Boston found a testing program to be "a very valuable instrument for maintaining standards in the Talmud Torahs" (49)

The Index to Shevile HaHinukh, Volumes 1-5 first series (1925-39) and volumes 1-18 (current series 1940-58) compiled by Zalmen Slesinger and Akiba ben Ezra, lists no articles on the subject. "The Jewish Teacher" in continuous publication since 1932, and "The Synagogue School," established in 1942, the educational publications of the reform and conservative movements in American Judaism also have **very** few references to the subject.

The only exceptions we could find are (a) The volume, "Testing the Knowledge of Jewish History" by Julius B. Maller, Union of American Hebrew Congregations (1932); (b) "Testing in American Jewish Education" by Zdanek Vanek, a thesis (typewritten) for the degree of M.H.L. at the Jewish Theological Seminar (1952), (c) "Jewish Education" periodical issued by the National Council for Jewish Education, which has many articles on the subject, and (d) Pedagogischer Bulletin (Yiddish), in publication since 1939, which has several articles on



testing and a large number of tests, mainly on Yiddish but also on holidays, Bible and other subjects. (e) Finally, is "Jewish Education in the United States" by Alexander M. Dushkin and Uriah Z. Engelman, published by the American Association for Jewish Education in 1959 (42) and containing the first national clarion call for the importance of testing and evaluation in Jewish education. It called for the organization of a National Curriculum Institute which is to do, among other things, "testing and evaluation of achievement and creating much needed instruments for the purpose," (p.175) It "should concern itself, as a major obligation, with the creating of standardized educational tests of all types and with using them in schools throughout the country continuously for evaluation and direction" (p. 250)

2 loose  
construction

b) National Organizations: The educational arms of the Orthodox and Conservative branches of American Judaism have done virtually no work in this area. The National Council for Torah Education, the orthodox national group working with Talmud Torahs, has "not done any psychological testing or evaluation of educational material in a formal way for many years...Many years ago (it) did evaluate tests that were presented...by various schools...as part of our daily work we do evaluate all kinds of tests that are included in manuscripts that are presented to the Vaad for publication." \* The United Synagogue Commission on Jewish Education, Conservative, "has done very little

---

\* From letter by the Director



in the area of testing." \* Reform Judaism had a Director of Research, Dr. Julius B. Maller, as early as 1930. In addition to the book on testing Jewish history (41), he prepared the Jewish history and Hebrew prognosis tests, the first and for several decades the only standardized tests in this area for the Jewish school. However, we find no information as to their use or on any further work in this area, except for some work by its New York branch, by this division in Judaism. These three divisions constitute the bulk of Jewish schools in the United States, other than the all-day schools. There are also three small groups where Yiddish constitutes a basic or the only language of instruction, The Workman's Circle Schools, The National Jewish Folk Schools and the Sholom Aleichem Schools. Their testing, however, is basically limited to New York City and is thus classified.

The only national organization which has and is continuing to work in this area is the American Association for Jewish Education. Late in starting, it was established in 1939, it issued its first series of tests in the fundamentals of Hebrew, on three levels, in 1954. In 1962 its National Curriculum Research Institute issued its Achievement Test in Jewish History. Both tests are widely used throughout the country, both in Bureaus and in individual schools.\*\* In addition to these standardized tests the Institute has also issued experimental tests on the High School level, in both Hebrew and English, on the Bible and on contemporary Jewish life, and a number of questionnaires.

The National Study of Jewish Education, conducted by the American Association for Jewish Education, administered achievement tests to several thousand

---

\* From letter by the Director

\*\* A letter from the Association indicates that in 1962 and 1963 it sold approximately 6000 history and 3,000 Hebrew test booklets, "Since these booklets are reusable, undoubtedly many more tests were actually administered." A recent statement lists 53 schools, 25 in New York City and 28 in 11 other communities. This however, is obviously incomplete, since reports from individual communities indicate at least 12 others using these tests.



students in nine communities in Hebrew, history, Judaism and Bible. It used the American Association for Jewish Education Hebrew tests and others without, however, preparing any achievement tests of its own which can be classified as standardized. (42)

c) Central Agencies of Jewish Education: The genesis of organized testing in Jewish schools has been traced by this writer to Boston in 1923, when candidates for graduation in the various schools were given a uniform test by the Bureau of Jewish Education, covering the work of the six year elementary school (47). This continued twice a year for 16 years, and in 1929 it was also introduced in grades 2 to 5, a program continuing except for an interruption between 1944 and 1947 to this day. For a number of years separate tests are given in the 5-day and the 3-day schools and in 1962 testing was also introduced in the Hebrew departments of the Reform schools. In 1962 over 4000 were administered.

Baltimore is the second and for many years was the only other school system to establish continuous annual city-wide tests. Established "about 1932,"\* they include "all major areas of instruction and are educationally sound, comprehensive, objective, reliable and locally valid." (48) They prepare their own tests in Siddur, Hebrew, H'mash, history and laws and customs and in recent years have also used the AAJE tests in Hebrew and history. In the late 40's and in the 50's it also administered Noah Nardi's Jewish Information Test. In 1948 it established a department of tests and measurement with Harry Tchack as part-time director, a department which continues to this day. About 2,000 children in 11 schools are tested annually. A feature of this testing is a local standard score system which they consider successful in interpreting the test results.

---

\* Report from Board of Education. An article in Jewish Education (48) gives the date as "about 1943."



The largest testing program is conducted by the Bureau of Jewish Education of Los Angeles, Begun in 1946 by Professor Morris Liebman for graduating classes, it was expanded since 1948 under the guidance of Dr. David Bridger in all areas and grade levels of the elementary Hebrew schools. Since 1962, the Bureau has also been administering a test in Jewish life and observances in the sixth and seventh grades of the one-day-a-week schools. Its Hebrew test is given to about 18,000 pupils in 60 schools annually. Smaller numbers of students take the tests in all other major areas of instruction.

Testing in New York City may be divided into two periods, beginning with about 1945 when Noah Nardi began preparing a number of standardized tests used in New York City and elsewhere and beginning with 1950 when the Metropolitan Council of United Synagogue (Conservative) instituted testing, administering some 10,000 tests in 3 years (50). For a number of years, the Jewish Education Committee has been preparing End of Third Year Achievement Tests for Weekday Afternoon Schools. In 1961 they were given in 76 schools and since then to about 100 schools. These tests are in Hebrew, Jewish Life and Observances and Jewish People. The JEC maintains some control by offering to score them by IBM (74). New York City does not have a permanent department of testing with a person devoting all his efforts to this work. A department of information, research and experimentation was established in 1949 under the direction of Dr. Israel S. Chipkin. In 1951 it began a survey of which one of the aims was "to evaluate the practices and achievements in the Jewish schools in New York City." (51) About 3,000 students were tested in this survey, about the same number as are now tested annually. Following the death of Dr. Chipkin the survey was completed by Louis Ruffman. In addition to the areas listed above it also included a home environment questionnaire



and one on attitudes of youth to elementary Jewish education. (77)

In a survey of testing in New York City we should also include a history test prepared by Max Nadel for the reform schools and the testing conducted by three Yiddish groups: The Sholem Aleichem, The Workmen's Circle and the Jewish Folk Schools. The first was prepared because of the negative reaction of principals in those schools to the AAJE history test. An experimental edition was administered to about 1000 students in grades 7, 8 and 9. This was followed by an item analysis and a final edition was prepared but "the principals of the Reform Synagogue School did not take readily" to this test either. He persuaded 15 schools to give the test and it was given to 323 students.\* The Sholem Aleichem Schools were among the very first to administer tests, starting as a teacher activity in the early 20's and beginning school wide tests in the early 30's. The Workmen's Circle began testing in 1944. The JEC gives central graduation exams to both school groups. Mid and end year tests are given in the Sholem Aleichem Schools and mid-year exams only in the Workmen's Circle Schools. The JEC supervisor considers the mid-year exams more important. The Jewish Folk Schools in New York City also give uniform tests in grades 4 and 5 in a central place in February and June of each year. The tests are in Yiddish but also in Humash, Hebrew and History. Detroit and St. Louis develop their own tests, the first beginning with 1959 and the second with 1960. The Detroit tests are given at the end of the Aleph class and twice a year thereafter. It places the students in their percentile rank and has a linguistic and non-linguistic median. A combination of both local and national tests

---

\* Letter from author



are administered in Minneapolis since 1945 and in New Haven since 1957. A very limited program is conducted in Chicago "practically during the entire period of the Board's existence" (organized in 1923) and a small program also exists in Cincinnati. Miami has been administering several thousand tests annually since 1950, preparing its own instruments in history, Humash, prayers and laws and customs. It also correlates the results of national tests it administers with those of several other communities. This process was begun by Louis Schwartzman in 1947, when in Atlanta, a program he transferred to and expanded in this community.

Full or limited use of national tests only are made by nine agencies in eight different communities. These are Atlanta, Buffalo, Camden, Cleveland, Newark, Philadelphia, Rochester and San Francisco. Almost all of these programs were begun in the late '50's and in the '60s. All of these administer the AAJE Hebrew tests and three also its history test. These Hebrew tests are also used in the six communities which combine national and local tests and the history tests in four of the agencies. Our survey indicates only four instances where central agencies used non-local tests other than the above, one used the Baltimore test, one from Los Angeles, one from New York and one used the New York State Hebrew Regents Examination.

In a survey of almost all central agencies of Jewish education, consisting of 25 of the 27 groups in all the 24 cities where such agencies exist, or a total of 93% of all such groups,\* (Outside information indicates that the

---

\* The author defines a central agency as an organization working with more than one school or with several branches of one school, and includes only those having full-time professional direction. Of the 40 groups listed in the AAJE Register 7 are actually single schools and 6 others have part-time or no professional direction.



remaining two groups have no testing program) there are 20 agencies conducting a testing program, or 80% of all respondents.

Table 3: Review of Testing in Central Agencies

	Number of Agencies	Number of Schools Tested	Number of Children Tested
No Testing	5		
Agency Tests	5	324	18,700
National Tests	9	73	7,900
Agency Tests and National	6	125	10,700
No Information	2		
	27	518	37,300

Eighty percent participation in testing among the central agencies for Jewish education in the country is a rather large percentage. It is, however, much less when we consider actual numbers and percentages of children tested. This is due basically to two factors: a) The small numbers tested in some agencies and the small number of schools some of them include, b) The grades and areas tested (of course, there are also some rather large communities which have no Bureau).

(1) Limited Testing: The great shrinkage in the numbers of children tested becomes immediately evident when we realize that in New York City, which has more schools than all the other central agencies combined, the number of schools tested is less than 10% and the number of students is considerably less than that. Small percentages are also evident in other communities. Thus, in Newark

---

\* The author defines a central agency as an organization working with more than one school or with several branches of one school, and includes only those having full-time professional direction. Of the 40 groups listed in the AAJE Register 7 are actually wingle schools and 6 others have part-time or no professional direction.



only 14 out of the 45 schools are tested, in Cleveland 7 out of 18 and in Buffalo 5 out of 9. A very small percentage of the Chicago schools are tested, even in Los Angeles and Miami 25% of the schools are not included in the testing and a maximum of 33 schools in Philadelphia is obviously incomplete. In some communities such as Baltimore which has one of the best testing programs in the country, and St. Louis, only a small number of schools are included in the Bureau. Only six central agencies include Reform schools in the testing, and then usually on a limited basis,\* and the bulk of the Reform schools in America are excluded from testing. Only 8 central agencies has more than approximately 20 schools each.

(2) Furthermore, and equally important, where tests are given, not all grades and not all areas are tested.

Grades Tested: Five central agencies give tests in all grades 1 through 6, four more in grades 1 through 5 and one in grades 1 through 4. In the Hebrew schools in New York City only grade three is tested (the Yiddish schools have more intensive testing). Two agencies test the graduating classes only and one begins with grade 3. Eleven agencies have some testing above grade 6, and of these, two test grades 7 and 8 and two more test all grades through 10. The rest test single grades through 9.

Subjects and Areas Tested: The only subject which is tested in all agencies administering testing is Hebrew, and here 14 use the AAJE tests either exclusively or together with local tests. Eleven agencies test Jewish history in whole or in part and of these 7 use the AAJE test, 5 of them exclusively.

---

\* Boston began Hebrew testing in Reform schools in 1962 and Los Angeles has just started a test on Laws and Observances in grades 6-7; others are Chicago, Miami, Rochester and San Francisco.



One of these agencies indicates it is discontinuing the AAJE History test, finding it unsatisfactory. A survey of subjects and areas used in all responding agencies indicates as follows:

Table 4: Agency Testing According to Subjects

	<u>Total</u>	<u>National Tests</u>	<u>Local Tests</u>
Total Agencies Replying	20		
Hebrew	20	14	8
History	11	7	6
Humash	7		7
Laws and Customs	7		7
Siddur	2		2
Current Events	1		1
Yiddish	1*		1
Intelligence Tests	2	2	
Attitude	1	1	
Behavior Reference Record	1	1	

\* 3 groups in New York City

On the secondary level, grades above 6, 11 agencies do some testing and of these 7 test only one subject. Three subjects are tested: 7 test Hebrew, 3 Laws and Customs and 4 History. Assuming this includes grades 7 to 10 and considering the number of subjects in the course of studies, this is obviously only a beginning and thus far a very small percentage.

A summary of testing in the elementary and basic level of study in the Jewish school in the central agencies of Jewish education can be obtained by classifying testing in the five major subjects in each grade. It is not



possible to know with complete certainty which subjects are taught in each grade- most agencies do not have a uniform course of studies for all their schools. It is safe, however, to estimate that Hebrew is taught in each grade, that Siddur can be tested beginning with grade 2, and taught through Grade 5, that Humash begins with grade 3, that history is taught for a period 3 years and a full course on laws and customs is given twice in the elementary grades. A review of the testing administered in the major subjects in each grade in schools of central agencies administering tests follows:

Table 5: Agency Testing According to Grades

	<u>Grade 1</u>	<u>Grade 2</u>	<u>Grade 3</u>	<u>Grade 4</u>	<u>Grade 5</u>	<u>Grade 6</u>	<u>Total</u>
Hebrew	6	9	14	12	15	7	63
History			3	5	5	3	16
Humash			1	2	2	2	7
Siddur		2	2	2	2		8
Laws and Customs		1	2	1	2	2	8
Total	6	12	22	22	26	14	102

This would indicate that in schools of central agencies administering tests about 50% test Hebrew in all elementary grades, 25% in History, 20% in Laws and Customs and only about 10% test Humash and Siddur, with an average of 25% for all elementary grades. Since only 80% do any testing at all this reduces the percentage to 22 and it is certainly considerably less in view of the fact that we don't know how many children in each grade are omitted from the testing for one reason or another (such as absence or non-inclusion of students because they are considerably below par or are in special classes). Then again some of the testing is not done each year in the same grades. This, of course, refers only to communities having central agencies for Jewish education.



Equally important is, what does the agency do with the tests? Unless the test results are carefully analyzed and the schools and teachers advised of the results and of the comparative analysis, there may be no value in the testing. Indeed one Bureau states that the results of its tests are on file but were not computed. Our information indicates that only Detroit and the Yiddish schools in New York City administer mid-year and end-year tests, and in fact the latter feel the basic value is in the mid-year test because it points out deficiencies and allows time for improvement, while end-year results don't always have as satisfactory a follow up because of different teachers and conditions. Comparisons with other school systems sometimes have the disadvantage of comparing unlike factors and the danger of trying to prove the superiority of one's own program

d) School Testing

Individual educators have undoubtedly prepared many instruments in an effort to gauge progress in achievement or to obtain information on other areas of learning: practice, attitudes and/or values, instruments which were used once or occasionally and then forgotten. The writer himself recalls preparing an attitude questionnaire for Jewish youth which was used at one time in a nearby community. These usually lack almost all requirements of a validated instrument and cannot be taken into consideration. We are, however, interested in learning the degree to which individual schools prepare or avail themselves of standardized tests which can help them determine on a comparative basis the progress of their schools from year to year, in knowledge, in practices and in values, and the degree of retention after the passage of time.



In addition to the more than 500 schools included in the survey of central agencies of Jewish Education, the writer has also made inquiries of a selected number of schools\* in non-Bureau communities as well as individual schools in New York City. Replies were received from 52 schools, including 34 schools in 25 communities in 16 different states and the remainder in New York City. The 34 schools had a total child population of over 19,000, with an average of about 600 pupils per school; 17 had over 500 each and of these several had over 1000. Only one had under 200 pupils. The New York schools were somewhat smaller, averaging 450, and here, too, almost half- 8 schools- had over 500 each with a total of about 8000 students and only one had below 200. Being large schools, almost all had full-time professional direction. From the standpoint of selection these, therefore, are schools more likely to have testing programs. It would also be assumed that those not having programs would not reply unless acquainted with the writer and this was indeed so in most cases.

Of the 34 schools, 12 indicated they do no testing. Of the remaining 22 schools, fully 21 did Hebrew testing, but only 8 tested history, 5 laws and customs and 4 Humash, with none testing Siddur. Sixteen used the AAJE Hebrew tests and 3 its history tests. When we break down the testing according to grades we find the following in the major subjects in the six elementary grades.

---

\* Selected by Dr. Judah Pilch



Table 6: Non-agency Testing According to Grades

	<u>Grade 1</u>	<u>Grade 2</u>	<u>Grade 3</u>	<u>Grade 4</u>	<u>Grade 5</u>	<u>Grade 6</u>	<u>Total</u>
Hebrew	5	7	12	13	12	6	55
History	1			2	2	2	7
Humash				1	1	1	3
Siddur				1	1	1	3
Laws and Customs	1			1	1	1	4
Total	7	7	12	17	16	10	69

The New York sample is not taken for a review of testing, as above, because of the 18 schools replying, testing is done in 16 schools and of these 14 take the JEC tests and are, therefore, included in the survey of central agencies. Of these 5 schools also give the AAJE tests. Because the JEC tests are given at the end of the third year and most schools do no other testing there is, therefore, a tendency for less overall testing in JEC schools reviewed than in others. On other matters the 52 schools can be taken as a unit. Of these, only 7 did testing of any kind above grade 6, one of them in New York City. These are divided as follows: Jewish history 2, Judaism 1, Hebrew 5.

Assuming the subject division in the different grades indicated previously, the comparison of the percentage tested between Bureau and non-Bureau schools is as follows:



Table 7: Comparison of Testing in Agency and Non-agency Schools

	<u>Bureaus</u>	<u>Non-Bureau Schools</u>
Hebrew	50%	40%
History	27%	10%
Humash	9%	3%
Siddur	8%	0
Laws and Customs	20%	10%
Average, all subj.	25%	15%

Of the schools reviewed, it is evident that on all scores the non-Bureau schools do considerably less testing than do the Bureaus. The percentage is undoubtedly much smaller in non-Bureau schools than appears above because the Bureaus are all-inclusive while the schools are selected and only large schools are listed.

Of the schools surveyed, the only one which has prepared an organized system of testing in any subject is Beth El in N. Belmore, whose educational director, Dr. Isaac Levitats, prepared a series of 24 objective tests in Jewish history based on the Pessin texts. There are four tests for each volume and there are parallel tests for each. They are the most extensive history tests prepared in several decades, should serve as a model for Bureaus and other schools to follow and with greater care for some testing standards- a brief critique is available elsewhere in this paper- it can be used nationally. Dr. Levitats has also issued some Hebrew and holiday tests including a novel feature, questions on attitudes and behavior in relation to the specific holiday.



The Park Synagogue in Cleveland headed by Dr. Leon Spotts, has "literally hundreds" in informal tests. These are combinations of objective and essay questions, for mid-year, end-year and other occasions. We saw 13 tests of similar nature prepared by the Emanuel Synagogue in Hartford. Temple Israel in Great Neck issued an interesting series of 18 tests in Hebrew, 3 for each of the first six grades divided into trimesters. The three tests for each year, almost all multiple choice, are on one sheet, and are given to the teacher in advance. The teacher may select questions to give to the students. Regrettably, the tests are unvocalized and poorly mimeographed and have some odd choice of questions. A school in Jacksonville has tests with answer sheets for the upper grades. Beth El in West Hartford prepared a test in Hebrew on an "original" theory that a child's language aptitude can be tested on the basis of two weeks of study of the particular language. It consists of 16 pictures to identify in Hebrew. Schools in Teaneck, Lynn, Cincinnati, Mexico City, St. Louis, Savannah and Chicago have also submitted tests prepared by teachers or principals. None is completely satisfactory, even disregarding the fact that no effort was made at standardization. In most cases these tests are not reused. An educational director in Rochester indicates he does extensive oral testing, to test the class rather than the students individually. The head of a fairly large New York school indicates he does no testing because he doesn't believe in it and another states that tests do not fairly evaluate progress. The head of a school in New Jersey says he never found a test to suit his program (nor



has he prepared any himself). One school giving the AAJE Hebrew test skipped a year because of a feeling that students may be acquainted with the test. One discontinued the JEC tests because it is "based on an idealistic course, which doesn't exist, at least in my school." We found school tests in Hebrew, history, Judaism and Humash, with one indicating regular annual testing in Siddur. One school prepared combined tests for several subjects, such as History, Bible and Current Events.

There is no indication that tests of one school are used in other schools in the country.

#### d) Classroom Tests

Classroom tests may be divided into two categories: a) those prepared by the teachers themselves and b) those prepared by others for the teacher's use. In the first category are the thousands upon thousands of tests prepared by teachers many times a year, each test used once and rarely seen thereafter. The second group may be further subdivided as follows: 1) Tests prepared by principals and by central agencies, not for review as in the case of mid-year or end-year tests, but as a service to the teacher to encourage the frequent use of tests, to assure that the tests will be well-prepared and thought out in advance, to train the children in the types of tests they will be getting at the end of the semester, to give the teacher a better and more objective idea of the success of his work and to allow him to make comparisons between one year and the following, when the same texts are used and classes taught. Thus the New Haven Bureau has dozens of tests in Humash and Hebrew, many of them prepared by teachers following specifications and guidance



by the central organization. These tests are prepared for every 2-3 chapters of the texts used. We note somewhat similar tests used in Minneapolis and Lynn, Mass.

(2) Tests issued by publishers: Many textbooks issued by the Hebrew Publishing Company, the United Synagogue, Behrman House and others have short tests at the end of chapters, sections or of the entire volume. This is so, for instance, in Highlights of Jewish History, Humashenu and in the workbooks to Haivri. Several publishers have issued separate tests, such as the Union of American Hebrew Congregations in its Soloff history series and Ktav in its Hebrew texts. These have some of the same purposes as tests issued by central agencies or principals.

(3) Published teacher-prepared tests for various occasions. Perhaps the only instance in this group is a series of tests on Jewish holidays published in the Jewish Teacher (78). These tests prepared in 1939-41, are on three or four different age levels, grades 2 to 9, to be given prior to the lesson to check the student's retention from previous learning. Although completely objective and fairly good tests- no effort was made at standardizations- we are not aware of their being reprinted or widely used. We also noted that a staff member of a Los Angeles school, in cooperation with others, consistently used teacher-made tests (79).

Classroom tests, in their various categories, differ from others only in their complete lack of efforts at standardization. They also differ in length from most other tests only in so far as they are short-term tests. Yet they are a good means for measuring actual achievement (97) in many cases.



However, classroom tests are also prepared at mid-and-end year especially since principals have not established the practice- and will do so even less as more national tests are prepared- to give tests to their classes on an organized, systematic basis. This is indeed, regrettable. As one teacher told us, giving tests is the best thing for the teacher-and, may we add, for the pupils as well.

Teachers generally are often unhappy about outside tests, whether prepared within the school or by an outside agency. At times it's simply the fear that their classes will not show up well in these tests. Sometimes it's a lack of confidence due to errors of judgment the tests purport to reveal. They place more strength on a child's classroom performance than on these tests. A teacher who knows a child well is often angered at the stress put on scores which she is certain are inadequate and invalid."(93)

Teacher-made tests differ from other classroom tests as follows: a) They are usually strictly one-time tests and even the teachers don't keep copies, b) Many are given orally or on blackboard and those mimeographed are usually handwritten or otherwise poorly done. c) They are rarely objective. While essay questions have their place, they must either be separated from the rest of the test or must be done in a manner to facilitate marking. d) They are rarely reviewed by other teachers and e) Their contents, especially in Hebrew are often based on retention of material whose retention is of no significance as the specific content of many a story is of no importance per se.

We suspect that the basic problem in teacher-made tests is that they are usually made in a hurry and "good teacher-made tests are never easily made in a hurry. They should begin with goals and blueprint and should involve much preparation in creation. Once made it can be used again..so it may eventually save



time...the proof of teacher effectiveness is in objective measurement made at the end of each teaching period...But assuredly, if no effort is made systematically and objectively to examine the individual pupil's mastery of the material presented, then to say that a...unit has been "evaluated" is to play havoc with word meanings and can only lead to dishonesty or confused thinking or both." (80)

## 2. Intelligence Tests

How widespread is the use of intelligence tests in the Jewish school? It is safe to say that it is virtually non-existent, except in the day school. The only central agency which does intelligence testing on an organized basis in all its schools is Baltimore, which has been doing this testing continuously since the inception of its Board of Education some 25 years ago. It administers the Otis primary test to students in grade 1 of its Hebrew schools, the Otis Intermediate test in grade 5 and the higher test to all new pupils in the Hebrew High School. Where there is a radical disagreement with the teacher's estimate the case is checked with the Hearing Revision of the Binet-Simon test.(48) The Board supplies the materials, staff members of individual schools administer the test and the Board checks the results, keeps records and issues information. It has found that while the public schools have cooperated on special occasions, it is difficult and highly impractical to use this source regularly. At the same time it met with no resistance or unpleasantness on part of parents, who on the contrary accepted it as a valuable asset.\*

---

\* From letter by Dr. Simon Bugatch



The Boston Bureau as recently as 1956 administered mental tests to all pupils in grade 1 where schools wanted to group pupils. These tests have been discontinued even as the homogeneous groupings are no longer practiced. The New Haven Bureau, following the Baltimore pattern, has introduced intelligence testing in the Aleph classes of Hebrew schools and for all entering students at the Hebrew High School. However, schools not having ability groupings find insufficient cause for administering the test.

In this connection it is worthwhile to note the existence since 1956 of a department of psychological service for Yeshivot at the Jewish Education Committee of New York. In 1960 the department served 15 schools with two full-time and two part-time psychologists. There is no indication of general testing but of serving the needs of individuals. (75) One psychologist states that "a large number of (Day) schools employ psychologists only for the purpose of admissions testing," that many use psychologists for "snob appeal" and that "the field is a most chaotic one." (76) One observer states that most of the modern Hebrew Day Schools he visited have their own psychologists who give I.Q. tests at admission and sometimes follow them up with achievement tests.\*

Of the 52 schools surveyed only 2 gave intelligence tests this past year and one of them, which had been doing intensive psychological testing, including Revised Stanford Binet and Rorschach closed down at the end of the year. Four others did testing sporadically, one giving the Otis test, one the Kuhlmann-Anderson, one a Stanford Binet and the fourth gave Reading Readiness tests.

---

\* Letter from Dr. Isidor Margolis 11/27/63



A small number of respondents made some reference to guidance and psychological services. Two educational directors simply indicated that they themselves are qualified guidance counsellors and a third stated that one of their supervisors happens to be one and they take advantage of his services. One refers individual students to a local agency, in another school the Jewish Family Service recommends professional help to families in need of it, one consults the public school psychologist in rare cases, one uses the voluntary services of professional psychologists who belong in the institution and one school uses a local psychologist to deal with disruptive children. Of the New York schools reporting only 1 indicated they make use of the JEC psychological service. Two other Bureaus indicate that some of the large schools use psychologists in special cases. The National Council for Torah Education had for a few years a Guidance Service to help the slow learner, but found little request for this service. It is evident that the organized use of psychologists and guidance counsellors does not exist in our afternoon schools and that even their sporadic occasional utilization is virtually non-existent. The situation is apparently no better in the Day School. One guidance counsellor says, "The need for guidance in the Yeshivos is critical." (81)

It is interesting that in this area where there is no problem of availability of good tests, where all the principal or supervisor must do is select and use one, few indeed take advantage of it. Why is there such an abnormally low number of schools using psychological testing and other services? Some clues may be gathered from replies to our questionnaires. One states they gave such tests but found them a complete waste of time because the school has ability divisions; another, that the principal who preceded him administered such a test to Aleph pupils to determine eligibility for an accelerated



second grade but found no real correlation between test results and classroom performance, so now they base acceleration on performance alone. One indicates that they used to give I.Q. tests but found they can get that information from the public school; another, that he encountered resistance to such testing from his Board of Directors. There seems little doubt that few schools recognize the need for this service partly because its basic advantage is seen, where it is at all seen, as an aid in grouping and grouping is not widely practiced either because the schools are too small to enable this division or for psychological or other reasons. The supplementary Jewish school is, it is felt, less in need of psychological services because the public school handles this area, because psychological problems often don't come to this school and because the Jewish school has the child for so comparatively few hours a week. The Jewish school is also a voluntary school and could, where desired, eliminate children presenting special problems from its class activities. Many schools don't have professional principals and when they do they are often not trained in this field and have an instinctive fear of psychological tests- fear of the unknown. They feel that this is an area reserved for specially trained people. Our principals, generally, also have a fear of the new and they are skeptics; "you have something, prove it to me"- and it takes time to prove it and it is so difficult to prove things in the field of education that its practitioners become content to leave things as they are and convince themselves that theirs is the best possible world.

3. Aptitude Tests- The problems in Hebrew language aptitude and prognosis tests are a) availability and b) validation. The problem is finding a test which will have a much higher correlation to actual achievement than does the



intelligence test. There is almost always a positive correlation between intelligence and Hebrew learning; the problem is the degree of correlation (45). In all we found six efforts at such tests which we shall describe briefly.

(1) The first recorded attempt to prognosticate success in the study of Hebrew was made from a battery of other tests by Dr. Evelyn Garfiel in 1928 (45) and carried out on some 70 children ages 7-12. It was administered at the end of the first semester of study in one Hebrew school and was a combination of Hebrew, intelligence, arithmetic, and a reading test. Despite the small number taking it, it was divided into two levels depending on the children's achievement in Hebrew. She found an overall correlation with achievement of .44 and on the Hebrew section alone the correlation was .63. The correlation with teacher's marks was .50. There is no indication that the test was validated, it had a rather limited sampling since it was not used again and to our knowledge it is not available.

(2) Hebrew Prognosis Test, forms A & B, by Julius B. Maller, 1929, was for high school and college students and therefore not relevant to Jewish schools which begin Hebrew studies no later than grade 3 or 4. (55)

(3) Hebrew Prognosis Test, forms C & D, for Elementary School Grades, by Julius B. Maller and Simon S. Silverman, 1931. This is the first published attempt at a prognosis test in Hebrew for elementary school children, stated to be for ages  $8\frac{1}{2}$  to 14 and elsewhere as beginning with age 10. It is too long, too complicated and in parts much too difficult so that the average child will not make an effort to complete it. It has over 200 questions and is heavily weighted with English vocabulary, some of it difficult even



for adults(ambergris, complot, ochre). There is no information on validation, the most important feature of any test. The manual states that the correlations are sufficiently high for prediction "as far as groups are concerned," but gives no details. Of the five parts, the first two are Hebrew reading, not comprehension and the last is English vocabulary intended as a test on general intelligence, so that the bulk is not prognosis in Hebrew per se. It doesn't account for the validity of individual items, nor for relative weights of each section, nor does it provide for reliable and representative norms. The test is not for 8 and 9 year olds, ages when children today by and large begin Hebrew School.

(4) Hebrew Reading Prognosis Test by Jerome L. Hershon, copyrighted 1960 is the newest and most complicated test, having 11 different parts with a total score of over 250. Although called "reading" it is both reading and comprehension. Part of it could be given before beginning the study of Hebrew and should, in fact, be too simple and hence useless to many when given as intended at the end of the year. It is intended partly for grouping, apparently in Hebrew, on the assumption that those who do poorly in the test are potentially poor in Hebrew. In contrast to the great detail in the manual on instructions, scoring and conversions to percentile ratings, there is absolutely no information on validation. The author simply thinks it will prognosticate Hebrew learning. Perhaps he's right! The test is divided into different abilities, visual, auditory, motor and language, and may help the teacher to spot areas which require remedial work. At this point it is not as yet a prognosis test. It is also much too late to wait a whole year to



begin prognosticating Hebrew learning.

(5) Hebrew Language Ability Test by Morton Rosen makes no claim to validation other than apparent group success in prognosticating Hebrew ability in his school and grouping students into ability levels. It is actually the simplest of achievement tests given after 2 weeks in the Hebrew School. He works on a theory that if a child cannot learn vocabulary quickly it is not worth while giving him a Hebrew oriented language course of studies. If he makes this the only criterion and if it is irreparable it may be dangerous from a learning standpoint. Those in the lower division, who were placed there by error, may never be able to be transferred. The test is simplicity itself; 25 words from which to pick the 16 corresponding to all pictures. There was no controlled experiment, no item analysis, no validity or reliability determination. Nor has he studied the degree of success of the selection. It may, as he says, be a good group test. Its simplicity is certainly no disadvantage. Having worked it for two years he should be able to determine some correlations. At present it is not, of course, a validated prognosis test.

(6) The Hebrew Aptitude Test by Noah Nardi, 1951, is certainly the best and most validated prognosis test. (86) It requires no Hebrew knowledge, can be used by children age 7 and over, is entirely based on Hebrew elements, is not too long nor too complicated. It determined reliability through the split-half method and obtained an overall reliability of .89. After 3 to 5 terms of study he obtained a correlation with actual marks of .58 in Hebrew school and .83 in public high school. This would be good if it were well validated. Actually, however, while several hundred were used in the experiment his correlation is based <sup>on</sup> only 56 students in Hebrew School



in an age group ranging from 7 to 12 and only 20 students in the public high school. He also states that the teachers' marks were of doubtful reliability and that the test may not have high accuracy in individual prediction but would be helpful in singling out the poorest and ablest students. One gets the feeling that the author was rushing into print, for an aptitude test requires a large sampling over several years. This test, no longer used, should be experimented with again to provide a more accurate test of its predictive ability. Barring that we have no standardized prognosis test of any kind for Hebrew on the elementary level.

For the present it is understandable that a Hebrew language prognosis test is rarely administered in the Jewish schools in America.

#### 4. Achievement Tests

Introduction: In order to help determine the nature of current testing in Jewish education, the writer has reviewed hundreds of tests obtained from the National Curriculum Research Institute, from Bureaus and from individual schools throughout the country and has chosen for review 168 tests. The vast majority were prepared between the years 1959 and 1963; others were undated but appear to be of recent vintage. The only tests included which date prior to the last decade were the history tests and the Hebrew prognosis tests prepared by Julius B. Maller and associates and published about 1930 but somehow omitted in Vanek's survey.(46) The tests chosen were not samples, random or otherwise, but were rather all the best available in the judgment of the writer. Reference will also be made to the Nardi tests, reviewed by Vanek, both because of their special significance and because almost all others he reviewed are listed, in a later form, in this survey. The tests selected are from 19 organizations; 2 national agencies, 11 Bureaus, 5 schools and one book publisher.



The tests included 2 prognosis tests, one attitude questionnaire and 165 achievement tests divided into the following categories:

Hebrew	84
Siddur	9
Jewish Life	19
Humash	27
History	23
Misc.	3

All but 12 are tests on the elementary level.

The preponderance of Hebrew tests in this survey is indicative of the emphasis of the American Jewish school. Hebrew is begun in grade one and usually continues throughout the years of the child's study in the Jewish school. In addition to being the major subject in the course of studies, the frequency of tests in this area may also be due to the comparative ease of noting progress in the subject, step #1 preceding step #2, while this need not be so in the English content subjects, and even in Humash- not being a graded text-one may begin at many different parts. There is a comparatively large number of Hebrew tests despite the great diversity of texts in Hebrew as compared with other subjects in the curriculum of the Jewish schools.

In almost all cases these tests, in so far as they are objective, emphasize recognition rather than recall. This is so of necessity, because questions of recall do not usually lend themselves to identical answers. Questions of recall also present the problem of marking partially correct answers and especially in language errors in spelling.

Many of the tests may serve a diagnostic purpose, to find weak spots in teaching and to evaluate relative merits of different methods or materials. Where, for instance, the test consistently shows considerable discrepancy between scores in sections stressing individual words and those testing the comprehension of sentences or the use of grammatical forms, there is an



indication of a need to shift the emphasis in another direction. Where the same or similar texts are used by the same or equally competent teachers in similar classes, but some use the direct method of teaching Hebrew and others the translation method or some stress conversation and others emphasize reading for comprehension, or some teach phonetic reading and others through the whole word method, the results of some of the tests when repeatedly administered may show the superiority of one method over another. Testing the results of teaching different Hebrew texts of similar vocabulary under similar circumstances may prove the superiority of one text over another. Thus the Chomsky test may be used to evaluate texts on which his word list is partly based. 12

Regrettably we find that the results sought in the various tests are almost invariably only a comparison of grades to determine promotion, or division into homogeneous groups, or to indicate comparative progress from year to year, or to compare grades with national or city norms. This is satisfactory and may serve as an encouragement in studies or give a school or a system a feeling of satisfaction, but it does not provide all the benefits which may be derived from the testing process. However, the tests in category (a) which follow may be used for diagnostic and other purposes.

a) Hebrew- Tests in the Hebrew language must of necessity basically be of a quantitative nature- the testing of abilities in the language, since the aim here is to measure facts, not ideas. In this sense testing in Hebrew is similar to testing in Siddur reading or to testing arithmetic in general education. As one noted educator says, "To be sure, mere achievement in terms of skills and abilities in Hebrew is inadequate. The concomitant attitudes, accompanying the teaching process, such as interest and desire to read Hebrew and to continue the study of its literature, are highly significant. Unfor-



Unfortunately, these concomitant attitudes cannot as yet be measured directly and objectively by any available techniques. But we can and should measure step by step the extent to which we approach our goal quantitatively and in terms of abilities."(108)

The Hebrew tests may be divided into several categories

(1) Comprehensive Tests On Several Years of Study

(a) The AAJE tests on Fundamentals of Hebrew, #1-3, are the most widely used Hebrew tests in this country. They were prepared in cooperation with a committee of well-known educators, are on three levels, have national norms and meet all the standards of internal validity, objectivity and practicability. Regrettably, these tests have apparently undergone only a partial item analysis, have not been correlated with an external criterion, do not have parallel forms and have not been tested for reliability. The levels create the problem as to which test is most suitable for a specific purpose and their reliability may be affected by a T-F section.(96) The overlapping grades are also confusing and there is the question if the method of selecting the vocabulary was proper, if, while the random sampling technique is satisfactory, it should not have been based on a composite of texts or vocabulary lists rather than readers. Also, the manual lacks all statistical information except for norms.

These tests have as their aim the testing of vocabulary, comprehension of sentences, comprehension of stories and recognition of grammatical forms in context, and seem to meet these aims well in the intermediate and upper level. In the lower level there is no section on sentences and it would seem preferable to have such a section rather than relying as it does on the stories



to meet this aim. The section on stories, generally, seems to lack validity because the answers require checking single words which may be found without understanding the story. We would also prefer an equal number of questions such as 50 or 100 to the present 75, 103, and 109, which appear somewhat confusing.

(b) The Chomsky Hebrew Achievement Test, #109-14, is certainly one of the very best tests in the Hebrew language both in the clarity of its objectives and in the careful manner of its execution. It is based on the "Basic Hebrew Word Lists for the First Three Years in Elementary Grades" of the author, which in turn was carefully compiled on the basis of frequency in the Bible and more specifically in the books of Breshit and Shemot, on Rieger's list, on common words found in most of 11 primers reviewed and on functional words. This is based on the author's conviction that the major aim of teaching Hebrew in the American Jewish elementary schools must be reading for comprehension and instilling a desire to read. He thus selected a 3-year word list to provide an adequate preparation for "an intelligent and an appreciative study of the narrative portions of the Bible as well as for the meaningful reading and appreciation" of stories of modern Jewish life.(95, p.55)

The current edition of the test, also on three levels, is a revised experimental edition and was recently administered to a broad sampling of afternoon schools in Philadelphia. As in the case of the AAJE test, this one also tests vocabulary, comprehension of sentences, and comprehension of stories, but the recognition of grammatical forms is tested only on level III. It is the only currently available test in Hebrew which has parallel forms to be used interchangeably and has many additional features: only 50 questions



(on two levels), equal weights, answer sheets, scoring keys, continuous numbers with examples in each section, short matching sections with unequal columns and questions in English on the stories, to test the understanding of the story even if they may not understand the questions in Hebrew. Somehow, level I has two stories serving the same purpose, instead of emphasizing another feature, such as recognition of grammatical forms. We would also eliminate 5 questions from level III to make it an equal 50 and two examples would be preferable to one. We would like to see the tests validated and tested for reliability and a manual added. Actually much of the information usually contained in a manual is found in the book, "Teaching Hebrew," and in articles in "Jewish Education." However, it would be helpful if this were provided separately and the statistical information and instructions for testing and grading added. This, in addition to testing the correlation between the parallel forms of the tests, is apparently the purpose of the current analysis of the tests in Philadelphia.

It should be noted that since this test is based on a specific aim and follows a specific word list, it may not be valid in testing other aims people may have in the teaching of Hebrew or in testing other texts or word lists. This is as it should be, although there may be a great correlation between teaching for different objectives. The only question one may raise in connection with the word list is, since it is partly based on textbooks, should it not be revised if many of the primers are no longer in frequent use?

The principle of comprehensiveness which the author follows carefully is most commendable. Almost every word in the list on which the test is based



is found in the test. However, the author apparently equates the use of each word, whether used in an active or passive manner in the test, whether it is essential for answering a question or not, even if it isn't necessary for the understanding of a sentence, or if its meaning can be guessed in context. Since the active, essential vocabulary in the test is comparatively small, it should follow a specific principle of selection. We are not aware that this is done.

We may also question the principle of using only words included in the word list of the particular year.(95,p.61). Certainly the passive use of words of earlier years seems in order. The author himself states that the learning must be cumulative, that "vocabulary learned at one stage must be reviewed at the subsequent stage or stages"(ibid.,p.56).

Two further comments: a) If reading for comprehension and pleasure is the purpose of teaching Hebrew, then one may perhaps question the inclusion in these tests of individual words in isolation, without relation to sentences which comprise the only basis for intelligent reading. b)The author indicates that "optional responses should be no less than four" (ibid.,p.61), but test level I, forms A and B, Section Iv-A and IV-B, have only six choices for 5 or 6 sentences, leaving several questions in each section with considerably less than 4 responses.

Assuming that this is the only acceptable word list for the study of Hebrew in the elementary grades and that the stated aims of teaching Hebrew are the only valid ones, then results in this test may also demonstrate relative merits in certain textbooks.

The author also feels that "the test results point to the conclusion



that the consistent use of the reading approach from the outset, with comparative elimination of mechanical reading, will yield considerable edge over methods stressing phonetics and oral exercises."(108) However, we fail to see the validity of the statement unless this were proven by controlled experiments.

(c) The JEC tests, #4-5B, are based on a word list of which more than 80% of the words are also found in the Chomsky Word Lists. It would, however, be difficult to establish a correlation between the two because here the word list is for one test while the Chomsky lists are for three tests. Here, too, the test consists of individual words, sentences, stories and grammatical forms, but these tests are more complicated (as far as we can tell: they change each year and are undated) both in the types of questions and in their multiplicity. It is based on actual practice in the better schools in New York. These tests meet the standards of objectivity and internal validity but they are not standardized. The item analysis itself was based on scores submitted by principals with only a sampling checked, rather than complete review by the organization. No effort was made at establishing reliability or norms, and the manner of selecting the words is not indicated. The individual tests for grades 1, 2 and 3 have been eliminated and replaced by an end of third year test. We compared forms B and C of these tests and found them to be variations, not parallel tests. We believe these tests would benefit by parallel forms, equal weights and fewer categories.

(d) The Los Angeles tests, #7-10 and 147-52, are rather similar to each other and can be considered as one, with the last 5, the most recent, used for our review. They are tests on grades 1-4. We selected #147 for Aleph and #152



for Daled as illustrations. They are good tests, short, clear, objective. The first has inadequate choices, three and even two, and tests only words and sentences. The true-false section has no advance information on guessing. The second test includes words, sentences and grammatical forms but has four different stories, too much of the same and the questions on the stories could also be in Hebrew. The tests would be clearer if the numbers were followed by letters rather than by other numbers. It also has a test on grades 1-3, which is a composite of three separate tests for these grades. Only one test, #7 First Year Elementary Hebrew, indicates it is based on a specific word list, not listed, and each fourth word chosen. Each of the other tests is based on several texts used in the community.

(a) The Baltimore test #138 is also completely objective and meets the standards of internal validity. It even has local norms. It has a careful scoring procedure. These tests are unique in having a city-wide local standard score system, helping in clarifying the interpretation of test results. However, section 1 has 42 questions, which is comparatively too many single words and of the simpler form, Hebrew words with English choices (and why the odd number?) Also, having two groups of numbers, one within each other, is confusing. One may also question the item analysis which leaves words like 1215, 112102 as meeting the standard of "success."

## (2) Other End-Year Tests

(a) Germantown test #136 was prepared to be given in all grades, but through Hay, and because of the wide range has words like "boy" and "cloud" in the same question. It has numerous errors in vocalization, very large matching questions and too many individual words.



(b) The North Bellmore tests #106, 107, are satisfactory

tests contentwise. The format and the pictures can stand improvement, they're too crowded, have too many divisions and are unnumbered. Test #107 should also have sentences and perhaps a story. The matching columns should not be equal.



Table 8: Standards Met In Selected Hebrew Tests

Table 8a- Analysis of Validity, Reliability and Norms in Selected Tests in Hebrew

Tests	Sets Aims and Meet Them	Sufficient Sampling	Well-Constructed	Clarity	Guessing Factor Adeq. Handled	Examples	Item Analysis	Professional Review	Correlation with Outside Criterion	Enough Items for Reliability	Well-Established Parallel Forms	Tested for Reliability	Norms
<u>Section (a)</u>													
AAJE #1-3	Yes	Yes	Yes	Yes	Yes	Yes	B	Yes	No	Yes	No	No	Yes
Chomsky Tests #109-114	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No	Yes	Yes	No	No
JEC #4-6b	Yes	Yes	Yes	Yes	Yes	Yes	B	Yes	No	Yes	No	No	No
LA #7-10, #147-52	No	No	Yes	Yes	Yes	No	B	No	No	Yes	No	No	No
Balto. #138	Yes	Yes	Yes	Yes	Yes	Yes	B	No	No	Yes	No	Yes	Local
<u>Section (b)</u>													
Germantown #136	No	No	Yes	Yes	Yes	Yes	No	No	No	Yes	No	No	No
N. Bellmore #106-7	No	No	Yes	Yes	Yes	No	B	No	No	Yes	No	No	No
New Haven #165-6	No	No	Yes	Yes	Yes	Yes	Yes	No	No	Yes	No	No	No
St. Louis #16-20	No	No	Yes	Yes	Yes	No	No	No	No	Yes	No	No	No
<u>Section (c)</u>													
Ktav #67-82	No	No	Yes	Yes	No	No	No	No	No	No	No	No	No
Savannah #14, 54-66	No	No	Yes	Yes	Yes	No	No	No	No	Yes	No	No	No
<u>Section (d)</u>													
Elsie Chomsky #11- Hebrew Comprehension 13	No	No	Yes	Yes	Yes	No	No	No	No	Yes	No	No	No
Chicago #142-45	No	No	Yes	Yes	Yes	No	No	No	No	Yes	No	No	No
N. Bellmore Oral Aural #105	No	No	Yes	Yes	Yes	Yes	No	No	No	Yes	No	No	No
<u>Section (e)</u>													
Boston #115-20	No	No	Yes	Yes	Yes	Yes	No	No	No	C	No	No	No
Detroit #121-24	No	No	Yes	Yes	Yes	Yes	No	No	No	C	No	No	No

Categories: A- very well; B- satisfactory; C- poor



Table 8b- Analysis of Objectivity and Practicability in Selected Tests in Hebrew

Tests	Always Scored Identically	Scoring Keys or IBM	Advance Inf. on Scoring	Adequate Manual	Equal Weights	Reasonable Length	Few Categories	Answer Sheets
<b>Section(a)</b>								
1. AAJE #1-3	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
2. Chomsky Tests #109-114	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
3. JEC #4-6b	Yes	Yes	Yes	Yes	No	Yes	No	Yes
4. LA #7-10, #147-52	Yes	Yes	No	No	Yes	Yes	Yes	Yes
5. Balto. #138	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
<b>Section(b)</b>								
1. Germantown #136	Yes	No	No	No	No	Yes	Yes	No
2. N. Bellmore #106-7	Yes	No	Yes	No	Yes	Yes	No	No
3. New Haven #165-6	Yes	No	Yes	No	Yes	Yes	Yes	Yes
4. St. Louis #16-20	Yes	No	No	No	Yes	Yes	Yes	No
<b>Section(c)</b>								
1. Ktav #67-82	Yes	No	Yes	No	No	Yes	Yes	No
2. Savannah #14, 54-66	Yes	No	Yes	No	No	Yes	Yes	No
<b>Section(d)</b>								
1. Elsie Chomsky #11-13 Hebrew Comprehension	Yes	No	Yes	No	Yes	Yes	Yes	No
2. Chicago #142-45	Yes	No	No	No	Yes	Yes	Yes	No
3. N. Bellmore Oral Aural #105	Yes	No	No	No	Yes	Yes	Yes	Yes
<b>Section(e)</b>								
1. Boston #115-20	Yes	No	Yes	No	Yes	Yes	Yes	No
2. Detroit #121-24	Yes	Yes	Yes	No	No	Yes	Yes	Yes



c) The New Haven tests #165-66, end year tests for Aleph and Bet, would benefit by a scoring key and answer sheets, #166 is too long but is intended to be broken down into two parallel tests, #165 has only 3 choices in several parts and would probably benefit by the inclusion of stories. An item analysis was taken, and will be used in the revised edition.

d) St. Louis tests #11-20 are poorly printed, with small type and not sufficiently vocalized. Tests #16 and #17 have long matching questions of equal columns and on vocabulary only. Tests #18-20 are not objective: They are translations of sentences to Hebrew or English, some very difficult, or answering questions of varying lengths. There is some odd Hebrew, including errors, and no indication of how to handle spelling.

### 3) Classroom Tests

a) Ktav should be commended for preparing tests in connection with a number of its texts, but they are far from satisfactory. While they are well printed and have drawings, they have complicated scoring, no indication of how to mark spelling, or changes in forms, have choices of two or even two pictures and two words. Test #71 has 10 words to supply the feminine; test #72, on the next 25 pages, has exactly the reverse. Tests #71-4 are simply inadequate not because of length but of content. Test #74 will illustrate: question I is spelling, II is singular and plural and III masculine and feminine. This leaves 8 pictures with two words between two pictures making half of them meaningless, leaving one other question to fill in missing words in each of 5 sentences.

b) The Savannah tests #14 and #54-66 represent part of an ambitious program of class tests and are, in fact, at times over-ambitious, as in test #14, when it has 73 questions in ten groups, including understanding a story, grammar, translation, sentence structure and much else- all to test 11 pages, or #62



with 54 questions on all of 8 pages. Nevertheless, structurally, they are rather good tests, except for long matching sections, unequal weights, equal numbers in both columns, insufficient choices. This is in sharp contrast to test #15 on all of Shalom Yeladim I which simply consists of 50 pictures with choice of 3 words in each. This test, however, is poorly mimeographed with badly done pictures, and may actually be inadequate to test the book.

#### 4) Special Tests

Elsie Chomsky's Hebrew Comprehension tests prepared some years ago for the Philadelphia Council on Jewish Education have but one aim- comprehension of stories- and in this they seem to succeed. In the first level, it would perhaps have been preferable to have the questions in English and in Multiple choice

The Chicago tests, #142-45, are especially designed for a specific program in Reform schools, and differ in content from all other tests. The vocabulary selected from the Siddur and Humash and the test is largely one of recognition of prayers and Biblical passages. The tests have some of the qualities of objective tests, are clear and well-constructed. The number of choices could be larger and the matching columns should be unequal.

North Bellmore test #105 is an attempt to test an aural-oral class after several months of study. It consists of three parts: picture identification, true and false and classifications, and seems like a completely acceptable procedure which is encouraging.

5) All-subject tests: Two communities, Boston and Detroit, give single tests for all the subjects taught in the class, the first as one unit and the second divided into separate sections.

Boston tests #115-120, end year tests for five-day schools, are the oldest



continuous annual tests in the country and have followed the same pattern for many years. This, by itself, is no disadvantage, as the pattern in no way gives advance indication of the content, and the content changes each year. These tests meet most of the standards of internal validity, objectivity and practicality. They are well-prepared and attractive. We did find several long matching sections, with equal columns, completions with only two choices and no advance information on guessing in True-False questions. The tests in the upper grades have essay questions which, in view of the large number of objective questions, adds to the test. Contentwise, however, the tests do not appear to test adequately areas other than Hebrew in all grades and Bible in grades 4 and above. There are, for instance, few questions and an inadequate selection on holidays in Aleph and only a handful of questions on holidays and history in Shemot and Daled.

b) The Detroit tests #121-24, have 4-5 tests in one and meet many of the standards of testing. We reviewed the mid-year tests for Bet-Hay and believe the first three have insufficient items for validity and reliability in any area, except in Hay where the addition of the Humash questions give greater validity and reliability to the Hebrew test. In construction it has many advantages: well-constructed items, examples, a scoring key, separate answer sheets and few categories.

#### (6) Secondary level

NCRI test #21 is a special test for an advanced High School group, based on a selection from Breinin's "Sefer Hasefarim." Though poorly mimeographed and unanalyzed the first five sections are objective and satisfactory. The remainder of the test consists of recall questions which should be satisfactory for this age group though it presents scoring problems which are not explained. The test has not been validated or tested for reliability.



b) Siddur: While the Siddur is perhaps the most used book in the Hebrew School we found only two tests on reading and three on content in addition to four tests on its related text, the Haggadah, a total of 9 tests. There are also questions on the Siddur in several tests on Laws and Customs.

1) Siddur Reading

(a) Group Test: Only one validated group test- in fact the only group test in Siddur reading was created by Noah Nardi in 1942(57). It was in the process of experimentation for four years. They were statistically analyzed and validated and tentative norms were created. However, reliability was established through retests of only 61 cases. This is especially low in this test because in half of it the correct answer is dictated and the variations in dictation and speech are certain to result in different degrees of accuracy at different times. The items were analyzed for success and discrimination but there is no information as to an analysis of the effectiveness of the distractors. Validity was established by comparison with teacher marks and individual tests but there are no details on the number of the latter. There is also no mention of the test being analyzed by teachers or specialists. In its present form it is not satisfactory; it can, however, be adapted to become a useful instrument, especially by the elimination of the dictation. Actually, questions 1 and 4 are identical, except that one is dictated and the other transliterated. The test seems to be too long and would probably be equally valid-in fact, more so, because it won't tax the patience of the student- with 50 as with 100 questions. Questions 3 and 4 should be shortened and retained and other questions found for one and three.

(b) An individual test in Siddur reading which has much merit was devised by



Berthold C. Friedl of the Chicago Board of Jewish Education and first administered in 1927-28 (58). It is based on syllables and may have validity despite shorter and longer syllables. Each child is asked to read a prescribed passage consisting of a given number of syllables previously determined. The examiner has a mimeographed reproduction of the text with space for recording information on the student; there is also a formula for computing reading index and space for listing types of errors, number of errors, time taken for reading and the reading index. He administered the test on 130 syllables from 312 7711 to over 2000 children in 11 schools. The formula followed was

$$\frac{(1 + e)}{S} t$$

e= number of errors                      s= syllables  
t= readingtime in seconds

Tests given by teachers themselves did not yield identical results apparently indicating that certain precautions must be taken to make the test useable. He suggests that experienced examiners retest a sample group of 2 or 3 from each class. There must also be definite standards as to what constitutes an error. Four types of errors were distinguished; pronunciation, omission, repetition and accentuation, but the last mentioned was not counted. He feels that the test tends to exaggerate the personality of the examiner and that the only solution is adequate training of the teacher in its administration. He gives the following average median index by semesters.

Semester	1	2	3	4	5	6	7	8	9
Index	7	4.64	3	2	1.3	1.8	0.9	0.9	0.5

This test as indicated in testing administered to thousands of pupils in grades 1-6 in Baltimore several decades later (59) shows a highly signif-



icant improvement from grade to grade, although Benathan (60) states that results of his study "indicate a low positive correlation between length of stay and proficiency in Siddur reading."

(c) Three individual Siddur reading tests were devised by the Baltimore Board of Jewish Education. Each test has 100 words in phrases, sentences or short paragraphs from the Siddur, clearly printed, and one test has practice lines. The instructor records the number of words read in 2 minutes or, if finished in less time, the number of seconds. An item analysis has been taken, it was tested for reliability, the scoring is satisfactory, it has a manual and there are city-wide norms.

The only questions I have are:

1) They count words rather than syllables so that a word of one or four syllables counts the same. This creates scoring inequity when the student does not complete the passage in the time given.

2) The student may often be acquainted with some of the passages, such as the blessings on foods and the timing of these selections would be a waste. I believe it would be preferable to select single words from the Siddur and to organize them in accordance with the number of syllables.

(d) In Temple Beth El in North Bellmore, the principal tests each child on a prayer he is to have learned that year and the score is the number of words in a minute less the number of errors. He has a frequency distribution chart, listing also the mean and median in each class. In New Haven the Bureau Direct- or tests individual pupils on an unknown passage and obtains two scores, one for correctness, deducting the number of errors in the specific passage, and one for speed.



2) Content: Of the tests on Siddur content, the Los Angeles test on prayers, #163, is well-constructed, has been analyzed for success, is basically objective and in parts practical. Contentwise, it's a good test but inadequate, with comparatively too much on the Amida and no reference to most other services. It is also an all-matching test with most sections having equal numbers in both columns.

Test #97 and 98 issued by the Minneapolis Talmud Torah are really one, with the second including all of the first. They are completely multiple choice but would do better with four choices instead of three. Test #97 has some errors in vocalization and deficiencies in mimeographing. A serious omission is the lack of questions on content or significance of specific prayers. Also, the title *Mishna* instead of Siddur just doesn't sound Jewish.

3) Haggadah Tests: The five tests, #91-95, may be taken as one. Their very existence is a virtue, as no one else has bothered preparing a test on the Haggadah. The tests are short and use few categories and the scoring is objective. Otherwise, however, they are deficient. Tests #91 and 93 have an English vocabulary attached, the mimeographing is poor and in parts unclear, have repetition of questions, choices of one or two, equal numbers in matching, have questions of recall and recognition thrown together, almost all questions are fill-ins or matching and have odd groupings of questions. No effort is made to test reliability, there is no validity, internal or otherwise, and it lacks most other requirements of an objective test.

4. Holidays and Observances: Of the 16 tests on Holidays and Observances on the elementary level, 8 of the 9 JEC tests are almost identical in form



and may be counted as one. The tests for grade 2 are the same as for grade 1, except for additions, and the same is true for grade 3. We compared forms B and C of the third year tests and found 32 questions out of 45 in Section I, all of Sections II and IV and 8 out of 12 in Section III identical. These are, therefore, not parallel tests- indeed they don't claim them to be. The three Los Angeles tests are also similar to each other. This creates a total of 7 categories. None of the tests was validated except for partial item analysis conducted in New York, Baltimore, Syracuse and Los Angeles. None was tested for reliability and only Baltimore has established city-wide norms. Six of the seven groups are more or less objective and four meet most of the criteria for practicability. None is apparently based on specific material, so that the degree of choice is difficult to determine.

Other comments on individual tests follow:

(1) JEC tests #25-32: The unequal weights have little validity- part 2 is partially easier than part 1 because of fewer choices for several questions, yet has double weights for each question; part 3 also has double weights, probably because it consists of prayers, but why is the recognition of a blessing on fruit more difficult than questions in part 1? In test #25 some questions have only 3 choices and in several places the same word appears both in the questions and the possible answers, giving it away. The tests have too few thought questions; test #32 has nothing on Purim; some names are given in Hebrew only when some children may know them in English, too few questions on ceremonies not connected with holidays (is there a boycott on Kashrut?)



- (2) JEC test #33 is a questionnaire not requiring a signature, has equal weights and good subject subdivision. It also includes 10 supplementary questions on home observance. Several questions such as *ענין יום כיפור* seem too difficult.
- (3) Baltimore test #139 has virtually no ceremonies other than holidays and comparatively too many questions on the Synagogue, but otherwise has good distribution, including questions on the content of prayers. On the one hand it has five choices which may be too many for elementary grades, and on the other it includes true- false questions. The groups are also divided into unequal numbers of questions.
- (4) The St. Louis test #104 consists of major holidays only; beyond that it is a combination of questions on Bible, history, Israel, geography and an essay. It is a most unobjective test, with all questions and answers of varying lengths and subdivisions.
- (5) Minneapolis test #99 is somewhat (too) long, has too many questions on the calendar, nothing on the content or significance of prayers and few on significance of the holidays and questions on numbers which are not directly related to the subject. A good idea are the questions requiring the use of a calendar.
- (6) Syracuse test (now North Bellmore) #130 is much too long-150 questions, and is confusing in the odd number of questions in the different groups. It has 11 matching sections-much too many, with equal columns, too many T-F questions and no special instructions. It is on holidays only and some odd content such as knowledge of words of some songs.
- (7) The Los Angeles tests #157-59 are all multiple-choice but have only 3 choices, #157 is on holidays only but somehow omits the Sabbath. It is poorly



balanced; 12 of the 50 questions are on Passover and yet its significance is omitted. Most are T-F with no note on guessing. Test #158 also has few questions on significance and no prayer content but has a better distribution and includes some non-holiday ceremonies. The test is (somewhat) too short for so wide a variety and its sephardic transliterations without the Hebrew may create problems in many schools. Test #159 is well-balanced and has many questions on content and significance.

We also reviewed three tests in this area on the secondary level.

(1) The Los Angeles Achievement test on Jewish holidays and observances for grades 6-8, #164, meets most standards. The Hebrew transliterations in Sephardic seem especially a problem since the test is apparently intended for Reform schools whose national organization uses Ashkenazic in its texts and programs. Actually, only 2 questions may identify the test with Reform schools and it can be used in other schools as well- although the omission of Kashrut, Tefillin and Tisha B'Av may also enter it in this classification; Generally it has a good sampling but few thought questions and why it has 85 questions we don't know. All the holidays are represented though Passover is emphasized with very little on Sukkoth. Several questions are odd or difficult. Is "the most important place for the observance of Jewish customs" the home or the Synagogue? Is the idea that "God has granted us the power to be better human beings?" suggested by the Mezuzah, Shofar, Menorah or Magen David?

(2) NCRI test #34 seems inadequate for its purpose, Its 50 questions are spread over the holidays, fasts and customs, with but one question on many subjects. The test assumes greater knowledge than is commonly taught in English-content classes, such as Ushpizin and Akdamut. Its expansion, especially with the addition of thought questions would make it a good test in this area.



Table 9a- Analysis of Validity, Reliability and Norms in Selected Tests in Holidays and Observances

	Set Aims and Meet Them	Sufficient Sampling	Well-Constructed	Clarity	Guessing Factor Adequately handled	Examples	Item Analysis	Professional Review	Correlation with Outside Criterion	Enough Items for Reliability	Well-established Parallel Forms	Tested for Reliability	Norms
<u>ELEMENTARY LEVEL</u>													
1. JEC #25-32	Yes	Yes	Yes	Yes	Yes	Yes	B	Yes	No	Yes	No	No	No
2. #33	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	No	Yes	No	No	No
3. Balto. #139	Yes	Yes	Yes	Yes	Yes	No	B	No	No	Yes	No	Yes	Local
4. St. Louis #104	No	No	Yes	Yes	Yes	No	No	No	No	Yes	No	No	No
5. Minn. #99	No	No	Yes	Yes	Yes	Yes	No	No	No	Yes	No	No	No
6. Syracuse #130	No	No	No	Yes	Yes	Yes	B	No	No	Yes	No	No	No
7. Los Angeles #157-9	No	No	Yes	Yes	Yes	No	B	No	No	Yes	No	No	No
<u>SECONDARY LEVEL</u>													
1. L.A. Jewish Holidays #164	No	No	Yes	Yes	Yes	No	B	No	No	Yes	No	No	No
2. NCRI Jewish Holidays and Observances #34	No	No	Yes	Yes	Yes	No	No	No	No	Yes	No	No	No
3. NCRI Questionnaire on Concept-Values #48	Yes	No	Yes	Yes	Yes	Yes	No	No	No	Yes	No	No	No



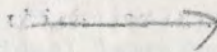
Table 9b- Analysis of Objectivity and Practicability in Selected Tests in Holidays and Observances

	Always Scored Identically	Scoring Key or IEM	Advance Inform- ation on Scoring	Adequate Manual	Equal Weights	Reasonable Length	Few categories	Answer Sheets
<u>ELEMENTARY LEVEL</u>								
1. JEC #25-32	Yes	Yes	No	Yes	No	Yes	Yes	Yes
2. #33	Yes	Yes	Yes	No	No	Yes	Yes	Yes
3. Balto. #139	Yes	No	No	No	Yes	Yes	Yes	Yes
4. St. Louis #104	No	No	Yes	No	No	Yes	Yes	No
5. Minn. #99	Yes	No	Yes	No	Yes	Yes	Yes	No
6. Syracuse #130	Yes	No	Yes	No	Yes	No	Yes	No
7. Los Angeles #157-9	Yes	Yes	No	No	Yes	Yes	Yes	Yes
<u>SECONDARY LEVEL</u>								
1. L.A. Jewish Holidays #164	Yes	Yes	No	No	Yes	Yes	Yes	Yes
2. NCRI Jewish Holidays and Observances #34	Yes	No	No	No	Yes	Yes	Yes	No
3. NCRI Questionnaire on Concept-Values #48	Yes	No	No	No	Yes	Yes	Yes	Yes



(3) Test #48, the NCRI information questionnaire on Value-Concepts is unique and worthwhile. It is a well-constructed test on a large variety of concepts which should be taught to our High School students and the bulk are thought questions relating to understanding and using these concepts. The many questions on their origin, however seem of comparatively little significance.

A basic problem in tests on laws and customs, it seems, are the items to be included and the degree of emphasis of each. Without reference to specific texts and extensive outlines it seems that they would of necessity be wandering all over the place, omitting and emphasizing subjects and areas in accordance with the momentary whim of the writer.

d. Bible: In the teaching of the Bible on the elementary level we found 25 tests from 8 communities, all in Hebrew and all on the Pentateuch, mostly on the book of Genesis. These may be divided into 8 groups in so far as types of tests are concerned. Contentwise, the 8 groups are all different, except for two tests on the book Humashi. Somehow, the JEC  which has many tests in other areas has not as yet made available any tests in the Humash; nor has the AAJE. The texts of Pollack, Scharfstein, Kaplan, Divinsky and Reiskin are represented. Only Baltimore includes content questions in English and New Haven does it partially. None of the tests has been reviewed for validity or reliability and only Baltimore has norms on a local basis.

There is some question as to what is essential in the study of Humash and hence what should be tested. It seems safe to say that this includes



a) Content, b) Translation of the Humash and c) Significance. This should, it seems, imply that a. and c. may be in English, in fact preferably should, except in outstanding Hebraically centered schools. It also means that many types of questions do not belong in a Humash test including a) translation from English to Hebrew in the language of the Humash; they'll never be expected to write a Chumash b) writing words in the language of the Humash c) translation of individual words from Hebrew to English except if we want to retain them in our Hebrew vocabulary and have no other way to test it. When classees have no other Hebrew instruction and use the Humash also as a language test, they should, in fact, give separate tests in the two areas, in Chumash and in Hebrew, d) spelling.

Special Comments follow:

(1) Test #24, issued by the Commission for the Study of Jewish Education in the United States, may have served its purpose of obtaining a general idea of Biblical knowledge or lack of it among our Hebrew school children. It is, however, too short and sketchy to be used as a gauge of learning at a specific time. Perhaps it's of some value in testing residual learning. Passing the Bereshit or Bamidbar test would probably indicate a knowledge of the books. This is not true of Shemot because the questions are on only part of the book or of Vayikra because only a knowledge of five words is required, even if in context. Failing any or all of the tests does not mean that the student does not know the Humash. It cannot be used in our Hebrew schools.



(2) Minneapolis tests #100-103 may be satisfactory for that school system but in almost all other communities our children do not, regrettably, complete the Five Books of Moses- they hardly finish the first book. The tests follow one pattern so that, except for content, reviewing one is reviewing them all. They all have adequate questions, few categories and of the multiple choice variety with 3 choices in each. Contentwise, they seem satisfactory except that in Bereshit half the questions are on Jacob and Joseph and only one is non-factual.

(3) Baltimore test #140 is the only one having a section on content in English. This test, however, is on the entire Humash and must of necessity be sketchy, even in 100 questions. Seventy questions are T-F, much too many, and it has no multiple choice questions. It has many single words out of context, many of which appear of little value in indicating a knowledge of the Humash, and there are few questions requiring reasons rather than facts. The test is based on the Kaplan series and would not be applicable to other texts.

(4) New Haven test #167, is the opposite contentwise, being based on only one Sidrah. It can, therefore, go into great detail and omit very little. Its vocabulary is largely in sentence context and it has a number of thought questions in Hebrew and English. After taking an item analysis it may be able to be split into two small parallel tests. If it proves too small then an effort at standardization should be made on two Sidrot rather than on one. Its scoring is most impractical without answer sheets which should be prepared in the future.



- (5) The Teaneck tests #134-35 are on 3-4 Sidrot each, are generally attractive and cover the subject fairly well. They give only 3 choices, are especially careless in vocalization and the mimeographing is in places unsatisfactory.
- (6) The Los Angeles tests #153-55 are on interrelated parts of the book of Bereshit, one on the whole book and two on parts of it in different texts. The distribution of the questions is good except that #154 concentrates on Abraham, Jacob and Joseph. They follow one pattern, are simple and clear but, though printed are poorly edited, each having a number of vocalization and other errors, in two tests at least 10 each. Test #156 is marked Shemot but is largely on two Sidrot, with nothing on most others. While mimeographed it seems more carefully edited and is otherwise similar to the others.
- (7) Ktav tests #83-90 lack most of the standards of good testing: the scoring is varied and confusing and could at times be reversed, equal matching columns, unnumbered questions, instructions in Hebrew only, two words with two choices, questions on spelling and numbers unrelated to the Humash. The tests are, however, well printed and each one is short, for several lessons only. They are based on the book Humashi which Ktav publishes.
- (8) The Savannah tests #125-28 are also on Humashi but do not follow the same subdivisions. For so little material the tests could be shorter and have less subdivisions. In the first two tests all instructions are in Hebrew-some of the instructions are obviously not understood-but that's changed in later tests. They also have unequal weights and equal matching columns. Contentwise, they seem satisfactory, though some English would help.



Table 10a- Analysis of Validity, Reliability and Norms in Selected Tests in Humash

	Set Aims and Meet Them	Sufficient Sampling	Well-Constructed	Clarity	Guessing Factor Adequately Handled	Examples	Item Analysis	Professional Review	Correlation with Outside Criterion	Enough Items for Reliability	Well-established Parallel Forms	Tested for Reliability	Norms
<u>ELEMENTARY LEVEL</u>													
1. AAJE #24	No	No	Yes	Yes	Yes	No	No	No	No	No	No	No	No
2. Minneapolis #100-3	No	No	Yes	Yes	Yes	No	No	No	No	Yes	No	No	No
3. Baltimore #140	No	No	Yes	Yes	Yes	Yes	B	No	No	No	No	Yes	Local
4. New Haven #167	No	No	Yes	Yes	Yes	Yes	No	No	No	Yes	No	No	No
5. Teaneck #134, 135	No	No	Yes	Yes	Yes	Yes	No	No	No	Yes	No	No	No
6. Los Angeles #153-56	No	No	Yes	Yes	Yes	No	B	No	No	Yes	No	No	No
7. Ktav #83-90	No	No	Yes	No	No	No	No	No	No	Yes	No	No	No
8. Savannah #125-283	No	No	Yes	Yes	Yes	No	No	No	No	Yes	No	No	No
<u>SECONDARY LEVEL</u>													
1. NCRI #35	No	No	Yes	No	Yes	No	No	No	No	Yes	No	No	No
2. #47	Yes	Yes	Yes	Yes	Yes	No	No	No	No	Yes	No	No	No



Table 10b- Analysis of Objectivity and Practicability in Selected Tests in Humash

	Always Scored Identically	Scoring Key or IBM	Advance Information on Scoring	Adequate Manual	Equal Weights	Reasonable Length	Few Categories	Answer Sheets
<u>ELEMENTARY LEVEL</u>								
1. AAJE #24	Yes	No	No	No	Yes	Yes	No	No
2. Minneapolis #100-3	Yes	No	Yes	No	Yes	Yes	Yes	Yes
3. Baltimore #140	Yes	No	No	No	Yes	Yes	Yes	Yes
4. New Haven #167	Yes	No	Yes	No	Yes	Yes	Yes	No
5. Teaneck #134, 135	Yes	No	Yes	No	Yes	Yes	Yes	No
6. Los Angeles #153-56	Yes	Yes	No	No	Yes	Yes	Yes	Yes
7. Ktav #83-90	Yes	No	Yes	No	No	Yes	Yes	No
8. Savannah #125-28	Yes	No	Yes	No	No	Yes	Yes	No
<u>SECONDARY LEVEL</u>								
1. NCRI #35	Yes	No	No	No	Yes	Yes	No	No
2. #47	Yes	No	No	No	Yes	Yes	Yes	No



We also reviewed two NCRI tests on the Bible on the secondary level, #35 and #47, the first in Hebrew and the second in English- the only English Bible test we found other than classroom tests. Both meet a number of test requirements. The first is on the Humash and latter prophets and appears to be a test on residual knowledge rather than on ordinary achievement test. It includes questions of recall as well as of recognition. As in another test by the AAJE, knowing the test indicates a knowledge of the subject but <sup>not</sup> knowing it does not necessarily mean a lack of knowledge. It has unequal groupings, varied types of questions with no indications as to scoring or spelling and too few choices.

Test #47 is not on content but on comprehension and as such seems both unique and good. It is a short test on a specific area and is, therefore, not in need of some normal requirements. The test is also experimental. Some of the questions seem rather difficult to answer especially since the student is unlikely to be acquainted with all of them, being on different parts of the Bible. Some of the English appears more difficult than the Hebrew.

e. History: The 21 history tests which we analyze herewith include 18 on the elementary and 3 on the secondary level and may be divided into 10 groups as indicated. These in turn belong to two sections a) five single tests on all of Jewish history and b) five tests on parts of Jewish history or on all of the subject divided into several tests. Two tests have been validated and meet all or almost all the criteria for good testing. It may be objected that the reviewers, while leading educators, were not necessarily experts in Jewish history. Reviewing questions of a test is insufficient. While the questions



individually may be worthwhile, they may not constitute together as a unit an adequate review of Jewish history. None of the others has also been tested for reliability and only the Baltimore test has local norms. The North Bellmore series has parallel tests on every part of Jewish history. The JEC tests (end of third year) are completely objective and meet the standards of internal validity but all the others also meet some or almost all of the standards for objectivity. All but #53 meet many of the standards for practicability.

Other comments on the tests follow:

(1) The NCRI tests #96 are certainly a model to follow in all areas of Jewish education. They are the first single standardized test on all of Jewish history and the only fully standardized tests in Jewish education now in use. In the view of this writer, however, they have several inadequacies: a) Fifty questions may not be enough to test all of Jewish history, b) The item analysis and tentative norms did not follow assurance that all who took the test studied all of Jewish history and that all learned it a more or less equal time prior to taking the test. Thus, one cannot compare results of one who completed the course this year and last year, c) one can get a passing mark without studying a considerable part of Jewish history or without understanding it. To help meet this objection it may be well to break up the scoring into divisions, d) We wonder if the final draft was reviewed by critics, especially for content validity for comprehensiveness after a large percentage of the questions were eliminated. Generally the distribution of questions seems satisfactory except that there are not enough thought



questions in form A, #4 and #15 are almost identical, #32 and #48 have the same answer and #2 is a poor question; 1/5 of the questions in each form are on Zionism and modern Israel but little on the medieval period or on the destruction of European Jewry.

(2) The UAHC tests are outdated because a) of the important period which followed their completion and b) they don't follow the order of texts now in use, have one test for all the past 2000 years, and two tests for the period ending with the destruction of the first Temple. The authors use a complicated scoring system, odd numbers of questions in various sections and long directions. These tests, by Julius B. Maller and others were, however, the most comprehensive and ambitious in the history of Jewish education in America, the only tests on which there is a full volume of information (41). Although issued by and for the Reform group, which has a fairly uniform course of studies, it nevertheless did not meet with great success. Perhaps the fact that the average pupil in the group tested scored only 36, with even less in test 1, scared them away. The test had an average reliability of .85, higher than well-accepted tests in general education.

(3) The JEC tests for the first three grades, #36-40, follow the same pattern as the tests in other areas it produces. They cover personalities throughout Jewish history and Jewish history to the end of the first Temple. This is in line with programs common in the Jewish schools but is a questionable combination. It would be preferable to have a test on personalities at the end of the second grade and then one or several tests on Jewish history in the succeeding years. The question on the geography of Modern Israel hardly belongs in this test.



Table 11a- Analysis of Validity, Reliability and Norms in Selected Tests in Jewish History

	Sets Aims and Meet Them	Sufficient Sampling	Well-constructed	Clarity	Guessing Factor Adequately handled	Examples	Item Analysis	Professional Review	Correlation with Outside Criterion	Enough Items for Reliability	Well-established Parallel forms	Tested for Reliability	Norms
1. NCRI #96	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
2. UAHC #41-44	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
3. JEC #36-40	Yes	Yes	Yes	Yes	Yes	Yes	B	Yes	No	Yes	C	No	No
4. #53	No	No	Yes	Yes	Yes	No	No	No	No	Yes	No	No	No
5. \$146	No	No	Yes	Yes	Yes	No	B	No	No	Yes	No	No	No
6. Baltimore #137	No	No	Yes	Yes	Yes	Yes	B	No	No	Yes	No	Yes	Local
7. Savannah #129	No	No	Yes	Yes	Yes	No	No	No	No	Yes	No	No	No
8. Teaneck #131-33	No	No	Yes	Yes	Yes	No	No	No	No	Yes	No	No	No
9. Los Angeles #160-62	No	No	Yes	Yes	Yes	No	No	No	No	Yes	No	No	No
10. North Bellmore #108	No	No	Yes	Yes	Yes	No	B	No	No	Yes	Yes	No	No



Table 11b- Analysis of Validity, Reliability and Norms in Selected Tests in Jewish History

	Always Scored Identically	Scoring Key or IBM	Advance In- formation on Scoring	Adequate Manual	Equal Weights	Reasonable Length	Few categories	Answer Sheets
1. NCRI #96	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
2. UAHC #41-44	Yes	Yes	No	Yes	No	Yes	Yes	No
3. JEC #36-40	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
4. #53	Yes	No	No	No	No	Yes	Yes	No
5. #146	Yes	No	No	No	Yes	Yes	Yes	Yes
6. Baltimore #137	Yes	No	Yes	No	Yes	Yes	Yes	Yes
7. Savannah #129	Yes	No	Yes	No	Yes	Yes	Yes	Yes
8. Teaneck #131-33	Yes	No	Yes	No	Yes	Yes	Yes	No
9. Los Angeles #160-62	Yes	Yes	No	No	Yes	Yes	Yes	Yes
10. North Bellmore #108	Yes	No	Yes	No	Yes	Yes	Yes	No



the distribution of personalities could also be better, since the bulk are through the first Temple, with 3 in the second, 5 others through Maimonides and no one else till Weizmann. Test #40, which we took as an illustration, also has 7 sections, too many for so short a test, and zigzags between one form and another.

A) JEC test #53, General Jewish Information Test, has 109 questions including on history. Contentwise, it is perhaps the best of the JEC history tests. It is all-encompassing, has many thought questions and the greatest concentration on modern times. It is, however, too weak on the early period, has too many subdivisions and they weave in and out. The series of questions asking to select the wrong statement in five seems rather good for retention.

B) JEC test #146 is a good test in that it is simple and easy to mark. As a test on all of Jewish history it is deficient in that it has nothing on recent history, Israel, or the US, except for a few questions on Zionism. The questions are basically in chronological order with a question on another period thrown in from time to time. Another shortcoming is the almost complete lack of thought questions. A number of questions seem unimportant. Several changes are also in order. Responsa are not just "letters," Marannos were not just "converted Jews from Spain during the Middle Ages," we don't classify Isaiah as a poet, and the question who was the founder of Islam doesn't belong in a test on Jewish history and we wouldn't say that he "was a prophet called Mohammed.")

C) The Baltimore history test #137 concentrates on the period ending with the Spanish Inquisition, having only 6 questions after that and none on Israel or the United States. It also has a preponderance of T-F questions.



- (7) The Savannah test #129 has nothing on recent Jewish history, few questions on Zionism and one comparatively unimportant question on the United States. It, too, follows the chronological order with a question on another period thrown in from time to time. It has too few thought questions and some of the constructions are wordy and clumsy. Also, five choices may be too many in elementary school tests.
- (8) The Teaneck tests #131-33 are good objective tests except that a) they omit some large areas: in test 1 there is virtually nothing after Solomon; in test 2 no Talmud or Gaonim, Rashi or Maimonides or even Spain except that they "were asked to leave;" in test 3 nothing on the United States or the destruction of Jewry; b) some sentences are poorly constructed, the question on historic sequence would be difficult enough to mark if it had 4 or 5 items, not 13, and 3 choices are insufficient in multiple choice questions, the questions on recognition of nearby civilizations are good, provided the textbook on which the test is based includes the information.
- (9) The Los Angeles tests #160-162 are on indefinite periods, not corresponding to known specific texts and two of the tests overlap considerably from the Maccabees to the Gaonic period, with virtually nothing before the Babylonian Exile. They have nothing on Medinat Israel. The tests are all multiple choice, which is good.
- (10) The North Bellmore tests #108 are undoubtedly the most ambitious and extensively organized group of tests in Jewish history. The tests are also short and completely objective. We reviewed test 1, form B, and found it has virtually no significant information on the period from the Patriarchs to the death of



Moses, except for Passover. Some questions are not too relevant, such as a short matching group including the date of the beginning of World War I and the Bolshevik revolution. It has about 40% T-F questions and equal matching columns. These tests should be reviewed and improved, tested for validity and reliability and published.

In addition to the above we received two tests in this area on the secondary level (tests #43,44 and 146 may also be classed in that category). The Information Questionnaire on the Jewish People, #45, is a well-constructed test, but has virtually nothing on recent or current history, no mention of Israel or <sup>Hitlerism</sup> Hitlerism, little on Palestine or the United States. Perhaps it's intended to be supplemented by test #46 on Contemporary Jewish life. This one, however, is of a lower quality, has many sections, not too well organized, having numerous areas of life, although almost one-half is on the United States and Israel.

6. Miscellaneous Tests: The three tests in this group are intended for Bar Mitzvah candidates. The first two are long and varied, having 200 questions each. The first is largely multiple choice with four choices while the third has only three. The second and third have large T-F sections with no instructions on scoring. The third also has large, equal matching columns. They are well-constructed, clearly issued and have enough items for reliability. The first two have virtually no features in the area of practicability. Contentwise, #49 has many faults. It is unevenly divided. In Humash it's actually only Bible stories and ends with Joseph, except for some quotations, and even then it's poorly balanced. In the Hebrew section it has 5 stories



Table 12a- Analysis of Validity, Reliability and Norms in Available Bar Mitzvah Tests

	Set Aims and Meet Them	Sufficient Sampling	Well-constructed	Clarity	Guessing Factor Adequately Handled	Examples	Item Analysis	Professional Review	Correlation with Outside Criterion	Enough Items for Reliability	Well established Parallel Forms	Tested for Reliability	Norms
1. Miami #49	No	No	Yes	Yes	Yes	No	No	No	No	Yes	No	No	ND
2. North Bellmore #50	No	No	Yes	Yes	No	Yes	No	No	No	Yes	No	No	ND
3. Los Angeles #51	No	No	Yes	Yes	No	No	No	No	No	Yes	No	No	ND

Table 12b- Analysis of Objectivity and Practicability in Available Bar Mitzvah Tests

	Always Scored Identically	Scoring Key or IBM	Advance Information on Scoring	Adequate Manual	Equal Weights	Reasonable Length	Few categories	Answer sheets
1. Miami #49	Yes	No	No	No	No	No	Yes	No
2. North Bellmore #50	Yes	No	Yes	No	No	No	Yes	No
3. Los Angeles #51	Yes	No	Yes	No	Yes	Yes	Yes	No



and 5 groups of questions when one or two would have sufficed and other areas could have been covered instead. The prayer section is devoted to reading only, without any effort at testing content or significance. Structurally, it is satisfactory.

#50 is a curious hodge-podge of questions of all types but largely on the synagogue and customs, a little on holidays and several questions thrown in on history and Bible. Both in content and structure it is a poor test and does not give a survey of Jewish content which a Bar Mitzvah test should cover.

#51, though short enough to be acceptable, is also not an organized test. It is largely on customs with a few historical facts added and not enough on holidays to test a knowledge of them effectively.

6. PRACTICES: Testing in practices, in Jewish living, is virtually non-existent in Jewish education today. As indicated previously, the check list and questionnaire can best be used in this area, but we know of no up to date instrument of a valid and comprehensive nature for this purpose. The changes in Jewish practices or the lack of them, over the past 30-40 years are so radical that, except as it applies to the comparatively small segment of orthodox Jews, an instrument of a generation or two ago can hardly be used today.

The earliest references we found in this area were a study by Jessie Ravich based on a questionnaire in 1926 and another, in 1928, by Jacob S. Golub. The first was administered to 389 men and women, ages 19 to 60, selected at random. The second, and much more comprehensive, was a survey of 1800 cases, ages 13 to 25, to determine the extent to which orthodox Jewish youth, affiliated with group agencies, practiced traditional customs and ceremonies. In its execution this



questionnaire obviously fell short in the variety of its sampling and it lacked scores or norms. Neither questionnaire is currently available. (109)

A questionnaire on Jewish home practices of students of Hebrew high schools and colleges was prepared by two educators and administered to about 600 students in 1956 (102). It is, of course, limited by its very nature, intended as it is for the small number attending these institutions. This, however, is only partly so, since 5 of the 7 questions can be used in a questionnaire with any group. These consist of two general questions on holiday observance and one each on the Sabbath, dietary laws and Synagogue attendance. Some of the questions are too general, such as "celebrating Jewish holidays through some home observance," or not sufficiently broken down, such as "abstain from work or travel on the Sabbath," or not listing some possibilities as in Synagogue attendance. There are also not enough questions for validity, unless administered to a very large number, and this is not possible with the group for which it is intended. The authors endeavor to establish an index of Jewish Home Environment (JHE) and set weights for each question. However, the weights are unequal and the reasons for the differences are not always clear. Why, for instance, should Synagogue attendance range from 3 to 0, home observance from 2 to 0, and holiday organizational observance from 1 to 0, or why should dietary laws be on a higher level than holiday observance. This is simply not a good questionnaire and almost all its questions are contained in greater and more satisfactory detail in the Nardi test to follow.

In 1957 the author of this paper prepared "A Survey of Jewish Youth," used with several hundred teenagers in orthodox, conservative and reform youth groups in a Jewish community of about 3,000 families. The questionnaire



is completely on matters specifically related to Jewish life and is a study of practices, attitudes and interests.

It has 19 questions on Jewish life and activity, 7 on Jewish interests and 8 on things the respondents think they would do when and if the occasion arises, such as inter-dating, intermarriage, give children a Jewish education, join a Synagogue and contribute to Jewish causes. It has a brief introduction intended to motivate the respondents to answer honestly. It has no place for a signature and allows place for explanations following some of the replies.

With some revisions the questionnaire could, I believe, give essential information on Jewish practices. It is less cumbersome than that of Nardi and is more up to date, but it should be tested for reliability, reviewed by other educators and a scoring key added. It should include the religious affiliation of the parents, be more specific on the respondents' Jewish education and add such practices as fasting on Yom Kippur, eating matzo on Passover and the lighting of Hanukkah candles. It also omits several other questions on the home: mezuzah and Jewish books. It should also state the sex of the respondent.

The questionnaire has two other sections which weaken it considerably: One is what is a good Jew, with 11 possible choices listed, and the other is five likes and dislikes, with open end answers. The first is difficult to score and the second proved to be a poor section, with few knowing how to answer or wanting to think of answers.

If revised the last part should be eliminated, it should be made entirely objective and the explanations should, therefore, not be included in the scoring.

A section on practices is included in the questionnaires for Hebrew and



Sunday School pupils which constitute part of the National Study of Jewish Education in the United States (42). It consists of 14 questions on holidays, 4 of these on home observance and the remainder relating to the students themselves. One of the latter may not apply to many students: not all children can participate in Purim carnivals. The questionnaire should, of course, list questions on other Jewish observances of the home and on practices of the children. The latter, could, for instance, include prayer and charity. The questions do include 3 choices and at times place for explanations. It has no weights. To be of use it would also have to be separated from the rest of the questionnaire and greatly expanded.

The only currently available instrument in the area of Jewish practices which bears consideration is the "Home Environment Test," by Noah Nardi, published by the Jewish Education Committee in 1946. It is a questionnaire consisting of two forms, one for pupils and one for parents/and family Jew-  
The test for pupils  
consists of 11 sections on personal  
ish observance, usually offering three choices (regularly, sometimes and never, or the like). The test for parents includes a similar section and adds one of opinion on Jewish education. After inquiring into the Jewish and secular background and related information, questions are asked on the extent of prayer, Synagogue attendance, Sabbath and holiday observance, kashrut, Yiddish conversation and Hebrew and Yiddish reading and participation in organizations and in charities. It includes a uniform scoring system and posits an index of Jewish Home Environment. The test was administered twice, one month apart, to 74 students, with a reliability correlation of .83.

The author conducted a study of 695 cases, children, adolescents and adults, and held some studies with limited groups on the relation between home environment and school work, Jewish information and teacher judgment of the



pupil's home environment. These studies were conducted with small and similar groups and were inconclusive even for the period in which they were studied. Even if completely valid and used in a large and varied population, such questionnaires must be repeated every five to ten years.

The Nardi test seems satisfactory both in format and content and, with proper revisions, could be used in this country at the present time. It should add religious group classification and synagogue membership. The section on Yiddish should, regrettably, be omitted, and the one on the Yiddish theatre is meaningless. The current test somehow omits fasting on Yom Kippur, parents attendance at Yizkor services and Kaddish as a Yahrzeit practice and could include attendance at services on Simhat Torah eve as a special category and waiting six hours after meat meals in the section on kashrut. The section on personal information requires some adjustment and the inclusion of the name of the respondent, unless optional, is unsatisfactory. I would place the child's own practices in a separate section and add to it attendance at public school on Jewish holidays and eating kosher food in the school cafeteria. This section should also be issued separately as a Jewish Practices Test.

The matter of length should not present an obstacle because the replies require checking answers only. While the questionnaire could be inclusive it should also suggest different rating for orthodox, conservative and reform groups, with some questions applicable only to a specific group. I believe it would also be a better test if it were less inclusive in listing information on almost all the members of the family. In its present form the Nardi test cannot be used.



A Jewish Home Environment Questionnaire was administered to 3115 students as a supplement to a test on holidays in the Survey of Jewish Education in Greater New York, 1951-52 (51). The questions are on synagogue attendance, kashrut, kiddush and candlelighting, five questions on cultural subjects and one on organizational activity. Almost all questions are applicable to all religious divisions in Jewry and the majority can apply even to secular Jews. It has an introduction to encourage honest replies and has no place for a signature.

The statement on services on Saturdays may be misleading since the respondent may not include Friday evening in this category; it should allow for partial replies: in addition to yes, no or I don't know, it should have sometimes; the question on taking the child to Jewish shows and concerts is hardly applicable outside of New York City and the one on taking the child to meetings is dependent on the age of the child- how many children below the age of adolescence would be taken to meetings: the question on hearing Jewish stories at home also does not appear valid except for very young children. The term Jewish club may be misleading: what is Jewish, for instance, in a Jewish country club? Even in so limited a questionnaire there ought to be something about Jewish holidays other than the Sabbath<sup>and on</sup>, Israel and charity. On the other hand, the inclusion of both kiddush and candle lighting on Friday evening is unnecessary- the first would almost certainly include the second.

This is too general and too sketchy a questionnaire to be of value except when used with very large numbers.

There are thus no satisfactory instruments for the testing of Jewish practices at the present time.



6. VALUES AND ATTITUDES:

This section deals with instruments for evaluating attitudes and values of students in Jewish schools and specifically gained in these schools.

While there are numerous adult and college youth studies of the character and attitudes of Jews (112), each follows <sup>ing</sup> its own pattern, there are no commonly accepted standards of Jewishness and they do not necessarily indicate the effect of Jewish education on the respondents.

In our survey, including all the ~~Bureaus~~ and numerous schools, we found only one community which indicated giving an attitude test\* and only two tests in this area given in individual schools on a limited bases. We found in all twelve instruments, mostly related to studies of Jewish school children or graduates or former students of Jewish schools. These may be divided into four categories: ethical behavior, attitudes toward the Jewish school, attitudes toward Jews and Judaism and theological concepts.

a) Ethical Behavior: The validity of these tests in evaluating the influence of Jewish education is, of course, open to question and is at best of limited nature because <sup>1</sup>a) the children when they come to Hebrew school already have characters, <sup>2</sup>b) they are under the influence of the Hebrew school a small percentage of their time, <sup>3</sup>c) much of the Jewish school program is not devoted to character training, <sup>4</sup>d) the Hebrew school subjects are largely unrelated to the child's major interests (92) and <sup>5</sup>e) character training is also the province of the secular school and the home, and <sup>character</sup>is also molded in the street and the other elements in the environment. Nevertheless, while we control only part of the time of the Jewish child, the teaching of values and

---

\* NCRI Attitude Inventory, Rochester, 1963



the molding of character must be one of our basic aims, and instruments of evaluation carefully prepared and widely administered may indicate differences between these and other children, if such differences indeed exist.

(1) A test for the evaluation of character for first-grade children was devised by Ruth M. Baylor (105). Intended to supplement other means of evaluation by the teacher, it consists of three parts: test one, "The Jewish Game," identifies 12 of 16 objects as Jewish; test two, "Happiness Game," presents three pairs of pictures, each showing a desirable and an undesirable activity; and test three, "Good, Better, Best", has four groups of pictures showing three different approaches to each, which the child has to decide as to comparative value. The first test consists of Jewish symbols as well as Bible scenes and the rest are general.

This test has merit but requires changes to make it effective: a) the elimination of part 1; a knowledge of facts of what is or is not Jewish, or even the identification of Bible stories as Jewish, is not part of values. An identification of what is Jewish does not necessarily imply identifying one's self as a Jew with the object. b) The test must be extended in two ways: (1) It is too brief, thus decreasing validity. It should have many times the questions it has, (2) it is basically limited to one value; equal treatment or sharing. The stories in the Five Books of Moses have so many values that it is surprising that the author concentrated on only one, even if it is for first graders.

Since <sup>do</sup> most of the questions <sup>deal with</sup> are not <sup>Hence</sup> on the Bible, ~~anyway~~, it would have been preferable to give them all general illustrations, based on principles depicted in the Bible stories. In that way <sup>the test</sup> it could be given before and after the year's studies to see if there <sup>is</sup> any change in the child's value



concepts. Of course, even then it will not determine if the changes, if any, are a result of the specific religious school studies or of other influences in the child's life. Incidentally, the test "Good, Better, Best" is a misnomer. It should be "Poor, Better, Good."

In any case, it is doubtful if the test can be changed so that it can be used to test values several years later.

(2) A Racial Attitudes Indicator, devised by Rose Zeligs, is based on Bogardus' "Social Distance Test" (103) \* This is intended to measure the social distance of children toward various races. Thirty-nine "races" were listed and the child could list seven choices of relationships: cousin, chum, roommate, playmate, neighbor, classmate and schoolmate. This was followed by intensive personal interviews of 12 children to ascertain attitudes and reasons. These children were again tested and interviewed at the age of 15 and 18 to note what changes have taken place. The first three were considered more intimate relationships, and the remainder more distant ones.

This test has little value except in determining distance between Jewish children and Jews on the one hand and all the other thirty-nine peoples put together on the other. We, therefore, fail to see the need for all these groupings. From the standpoint of testing brotherhood we fail to see the difference of disliking a Japanese, a Negro or a white American and proximity to and some knowledge of each group undoubtedly is a major factor in the attitudes, rather than training alone which would measure the results of Jewish schooling. Also, we don't know if we want to train our children to be chums and roommates with non-Jews- relationships which on an older level could be equated with inter-

---

\* E. S. Bogardus, "Immigration and Race Attitudes," D.C. Heath & Co., N. Y., 1928.



marriage- whether the Jewish school wants to encourage intimate relationships with other groups. Also, does a more distant relationship with other peoples mean less brotherhood?

From the standpoint of the Jewish school the relationship to all these national groups is meaningless, since the children are not taught distinctions among all these groups. If we limit it to one group, say Negroes, we would have to set up a different criterion and different values; from a practical viewpoint it would also be best to include other values, rather than limit it to one area.

On the scale of values it is also a poor test: a) how does a chum differ from a playmate and why is the first in the intimate group and the second in the distant one? b) How do you differentiate between the closeness of a neighbor and a classmate etc. and even if we could, can a child do it? It would be preferable to make seven categories, A to G, and differentiate between attitudes on that basis.

As indicated in the editor's brief introduction, the Indicator also lacks information on home background of the respondents. It should be added that information is also essential on Jewish schooling- the study did not even consider distinctions between children who received a Jewish education and those who did not.

As a study the sampling was insufficient and the manner of selection not indicated. Generally, how does one distinguish that part which was acquired in the Jewish school, except by large samplings of carefully selected people who attended a Jewish school for a number of years and other groups which did not?

It appears that the test was not used beyond the one-time effort of the author, and rightly so. It is a poor study.



(3) A "Questionnaire on Pupils Interests and Attitudes" (#52), prepared by the National Curriculum Research Institute, is basically on general values not related necessarily to Judaism or to Judaism alone. It is obviously intended for high school age and above and consists of three parts: 25 True-false questions, 100 words to which one is to react agreeably or disagreeably, and 20 questions almost entirely based on Jewish sources which one is asked to accept or reject.

The first section has many questions which are so obvious, it seems to me, that one will be strongly tempted to answer them in the manner he feels is expected of him. Furthermore, the answers to some of the questions are only partly true or are subject to disagreement: Are Americans better than Germans? After Hitlerism this seems hardly a fair question to ask of a Jew. Are boys and girls who can't talk Hebrew poorly educated? That depends on definitions; from a Jewish ideal, I would say, "yes." Do people who do not attend services have a religious outlook? Many would say that the converse is true, that people with a religious outlook are likely to attend services. Why is Israel sending technicians to Africa? Whatever the reason it couldn't do it if it were not comparatively superior in a technical sense. Together, the validity of the entire section is very questionable.

The second section has but two possibilities, agreeable or disagreeable. Actually, many of them may get neither reaction, assuming the person understands all the terms. Also, the references to many of the terms are of questionable value concepts. Really, what reaction is England, Confucius, Geneva, Septuaginta, Olympics or Mandates supposed to evoke? How does one interpret positive or negative reactions to such terms?

The bulk of the questionnaire is on general values- peace, democracy,



patriotism, justice, brotherhood, equality, liberalism, with little reference or necessarily relationship to Jewish sources, learning or influences. Is it the assumption that liberalism, for instance, is a product of Jewish education or that the questionnaire will help us ascertain if this is so? Or is the purpose of the question merely to find out how liberal the respondents are? Either way, it seems like an inconclusive test.

Section III also has some questions where the desired answers are obvious; several are rather difficult to answer in one word and others are not easy to answer, certainly not by high school students.

I believe it would be preferable to limit the test to a much smaller number of concepts, have less overall items, indicate the aims of the questionnaire and have each item correspond to the aims, and also not rely, as he occasionally does now, on one question per idea. Especially in so large a questionnaire it would be well to set weights to each item, preferably uniform at least in each section.

A major question, of course, is what value the test has, even assuming it is changed as suggested. I assume its purpose is to find out the reaction of young people to numerous ideas, values and attitudes, but this is not necessarily related to learning and ideas acquired in the Jewish school.

(b) Attitude Toward the Religious School: The pupil's reaction to his school, the teachers and the program of studies, undoubtedly affect the success of his studies and may have lasting results in his attitude toward Jewish life and Judaism. As one educator said:"...it rarely, if ever, seems appropriate to study progress in Hebrew language or in prayer and worship studies without explicitly relating such progress to the attitudes toward the Jewish school....," among other things.(99) We should, of course, study



the correlation between a positive attitude to Hebrew school and to Judaism.

In this area we found only four instruments:

(1) The most exhaustive instrument in any of the categories in this area is Noah Nardi's "Hebrew School Attitude Test," forms A and B. Each form consists of 35 different reactions to Hebrew School, from the most favorable to the most objectionable. It is the only attitude test having two forms intended to be given at the beginning and end of the year, but this is actually unnecessary. The attitude may be largely a result of the pupil's relation to his teacher. It would, therefore, be best to wait two years to repeat the test and after so long a period the identical form can be safely administered.

At least six, perhaps nine statements in form B cannot actually be marked. "There's always someone in the house to help me with my Hebrew homework"- perhaps there really isn't?; "Our Hebrew School is not far away from home"- perhaps it really is? Similar problems exist with #17, 21, 22, 24 and perhaps also with 1, 8 and 10, Form A has fewer objectionable statements: only 17, 21, 23, 24 are not necessarily related to attitude and some of the other statements are also not error proof, in so far as attitudes are concerned. While it's generally intended for both boys and girls, there's one question on becoming a rabbi and another on Bar Mitzvah.

I also question his scoring system: giving a different score for each of 35 questions must of necessity be subject to question. Even the statement, "I hate to go to Hebrew School," gets some credit. He would do better to place them, for purposes of credit, into several groups, as he does in the case of norms. However, the norms too, are subject to question. In Form A why are #15, "Hebrew School has its good points," and #17, "Many of my friends



go to Hebrew School," neutral. The only really neutral question in this form is #20, "I haven't any definite like or dislike," and this he lists as "mildly opposed." There are many others which appear objectionable.

On the other hand he could add: "I like/don't like to go to Hebrew School because of the teacher, the principal; I find Hebrew School easy or difficult." He could add reactions to different subjects or activities. He should also ask how long the student is in the school, not merely which grade he's in- it may make a difference.

In sum, despite its scientific veneer- norms, scoring, parallel forms, manual and article, it is not a satisfactory test. It is, however, a first test in this area and could be of some value if the suggested adjustments are made- and it can just as well be shorter than it is.

A questionnaire which appears to be an abbreviated form of the Nardi test is included in the Survey of Jewish Education in Greater New York, 1951-52 (51). We could not see a copy but from the report it follows that it has 17 questions of the same nature and includes a similar scoring system.

(2) A very brief questionnaire on attitudes toward Jewish education was prepared by a Cincinnati Reform Rabbi and used with a small group in his congregation (101). It contains only 6 questions and two of them are on educational background. The four questions on attitudes are on memories of, feelings toward and the effect on the individual of Jewish education and differences, if any, between those exposed to Jewish education and those not. Each question has 3 choices, and they are simple, clear and do not require a signature. In listing background it does not distinguish between different types of Jewish education. The questions on attitudes resulting from Jewish education are,



of course, very limited but if used with large numbers it should have considerable validity.

(3) A "Hebrew School Questionnaire," which probes attitudes toward the Jewish school and various facets related to it of a temporary and tentative nature was devised and used with some benefit by Anne Henkin. It contains six questions, with three choices and a "why" in most of them: do you like Hebrew School, which subject do you like best, do you like the class, is there anything you'd like changed, would you like a change in teachers next year and do you feel you made progress in your studies this year. It is anonymous and the introduction further helps in obtaining objectivity.

Few teachers would have the strength of character and self-confidence to use it, as they must be able to accept criticism. It could, of course, be given by the principal but it may be interpreted as encouraging criticism unless it's known that the teachers approve of it. The title is a misnomer, as it can be used in any Jewish school. With proper motivation it can represent a valid criterion of the reactions of a class to these questions and it could be of benefit to the teacher and the school in relation to these limited questions.

(4) The beginnings of a Jewish Appreciation Test (JAT), along the lines of Murray's "Thematic Appreciation Test," were prepared by Joshua A. Fishman in 1954 (50). He prepared picture plates dealing with problems revolving about the Jewish school, including attitudes toward the studies and the teachers, as well as problems of homework and conflicts between Christian or general customs and activities and Jewish ones. Children beginning with age 8 and 9 were invited to make up stories in connection with each picture- and they were very revealing. The author stated that the plates were being revised



but there is no indication that this was accomplished. There is also no indication that it was ever evaluated or used on a large scale. It requires the administration and analysis of a trained psychologist and <sup>is</sup> thus of little practical use in the current Jewish school.

(5) A questionnaire on attitudes and behavior constitutes a supplement to a Test on Tu Bishvat prepared by Cong. Beth El in North Bellmore. The first section includes attitudes on Tu Bishvat, Israel, Siddur, Hebrew School and Jewish history- of a varied nature but things the children like to do right now. On this basis, too, it's <sup>so</sup> incomplete: why seek attitudes on history and not on many other areas of study? The question if the child likes the fruit they give him is a poor one; perhaps it wasn't worth liking. Only two choices are inadequate for the question, "Do you like Hebrew School"? and this includes a number of others. One also wonders as to <sup>the</sup> connection between Tu Bishvat and almost all the other questions. Most of the questions should be included in a separate instrument and the selection is a rather poor one.

The test is too brief and too obvious; many children will answer "yes" to most questions because they are sure you want them to, especially since the name of the student is listed.

(6) Two questionnaires, one for Hebrew School and one for Sunday school pupils, but almost identical, constitute part of the National Study of Jewish Education in the United States (42), administered to over 10,000 students. It includes questions on attitudes toward the Jewish school and the public school, reactions toward individual subjects and selected questions on holiday practices.

While the questions comparing attitudes toward the Jewish and the secular school are helpful, placing them together may affect the accuracy of the replies.



It would be preferable to have separate sections on each. The requested evaluation of a long list of subjects is confusing; half the list would be preferable. The questionnaire serves the purpose of the study but is of little value for general use.

(7) Two questionnaires on attitudes of Jewish youth to elementary Jewish education were part of the Survey of Jewish Education in Greater New York, 1951-52 (51).

a) The first is for children who had some Jewish education. It includes information on characteristics of the respondents and then lists eight questions on attitudes, providing from three to fifteen responses to choose from. The questions are on attitude toward the Jewish school, the importance of Jewish education, comparisons with public school on interest and importance, reasons for drop-outs, benefits derived, most and least liked subjects and most and least worthwhile features of the school.

We were not able to obtain a copy of the questionnaire but from the survey it appears to be satisfactory in every way for the limited area studied.

b) The questionnaire for those who did not have a Jewish education includes four questions: reasons for not attending, feelings about non-attendance, attitudes toward Jewish schooling and subjects they would study if they had the time. For this group, especially, I would add questions on attitudes toward Jewish values: Israel, Jewish unity, charity to Jewish causes. Nevertheless, sight unseen it appears satisfactory for further use.

(c) Jewish Values: It is rather surprising that those values which can be largely attributed to the Jewish school training are almost completely absent in the available instruments in this area: a feeling of kinship toward the Jewish people, the land of Israel, the Hebrew language, a desire for



identification with and emulation of our great leaders, and a reverence toward Jewish traditions, an appreciation of Jewish practices, and a positive attitude toward Jewish study. In addition to some individual questions in 1c and 2e above and in this writer's "A Survey of Jewish Youth," contained in the section on practices, we find only one questionnaire on one phase of this question, identification with Jewish life. This is "A Questionnaire on the Correlation Between Elementary Jewish Education and Interest in Jewish Life and Problems," devised by Samuel M. Blumenfield (104). The questions were the amount of Jewish schooling, if any, and the degree of participation in Jewish organizations, if any. The first lists four categories: Hebrew School more than three years and less than three years, Sunday School and none; the second is divided into three groups: leaders- those holding an "important" office; active-those affiliated; and passive-those unaffiliated. The questionnaires were sent to young people who attended Jewish schools and to those who belonged to Jewish organizations. The first aimed to find out what proportion of those who had studied was active in Jewish organizations and the second to discover what proportion of the affiliated had a Jewish education.

The questionnaire suffers from a number of inadequacies and weaknesses:

(1) Definitions and classifications : (a) can less than three years of Hebrew School- it may apparently be even one year or less- be considered Jewish education? Perhaps I should say, can less than five years of Hebrew School be considered Jewish education? If less than six years of public school is placed by many in the category of illiteracy, can we actually say, as this implies, that any amount of learning, no matter how little, is Jewish education? (b) Agreeing that Sunday School is Jewish education, can



this be so indiscriminately, without reference to the number of years of study? Certainly one or two years cannot be placed in this category. It's rather surprising that in the case of Hebrew School he lists two categories, but he makes no distinctions between degrees of Sunday School attendance.

(c) He defines leaders as those holding an important office but doesn't state which offices are considered important. (It is possible that this is classified in the questionnaire, which is unavailable, but omitted in the article.)

(2) Assumptions: (a) The author assumes that participation in any type of Jewish organization is indicative of interest in Jewish life and problems, but this certainly need not be so in belonging to philanthropic or to social-cultural organizations. These need have little or nothing of Jewish interest in them and one may join them because of a desire to socialize or in the case of philanthropic groups because he has a "good heart." (b) He equates participation in Jewish organizations with an interest in Jewish life, thus assuming that non-participation is equal to non-interest. This is obviously false. Jewish organizations are far from all of Jewish life. In fact, an interest in more important phases of Jewish life, such as study, may preclude belonging to an organization on the part of some young people. Regular synagogue attendance, the practice of laws and customs, intensive reading in Jewish sources, travel to and living in Israel may show more interest in Jewish life than belonging to Jewish organizations. Isn't attending a Hebrew High School or College at least as significant as belonging to a Jewish organization? One may simply not be a joiner or be busy with other activities. (c) Even the search for a correlation between Jewish



education and leadership may be invalid. Isn't leadership ability a quality by itself? We want to train our children for participation in Jewish life and then if they have the ability they will lead. It is true, though, that greater interest may also be a factor in one's assuming leadership. Also, as Dr. Chipkin said, perhaps there is a correlation between a high I.Q. and leadership? (d) The author assumes all ages from 12 to 25 and over to be on one level but certainly the participation of the teenager, the college student and the young marrieds cannot be equated. (e) There is the further assumption that mere attendance at a school implies a Jewish education. There is <sup>not</sup> even ~~no~~ verification of the attendance or if it were regular, let alone if the people tested were diligent in their study as indicated by marks or statements by teachers.

The study would gain validity if (1) it had only two categories, participating and non-participating (2) it expanded the categories and eliminated groups of a more general nature (3) it limited the ages and (4) it defined Jewish education more carefully and, at this time, also added Day Schools and higher Jewish learning.

As conducted, again to quote Dr. Chipkin, this study has "serious limitations" and the conclusions are "tentative and open to question."

d) Theological Concepts: A "Religious Ideas Test" was devised by Abraham N. Franzblau as part of a study of Jewish adolescents (110). It contains 13 questions: 9 on the meaning and attributes of God, 2 on the meaning and efficacy of prayer, one on the meaning of religion and one on the immortality of the soul.

It is quite evident that a test of this nature, being as it is on theological rather than spiritual and ethical concepts, has no place for youth



averaging 13 years of age to whom it was administered. Children of that age rarely learn these ideas--and superficial learning isn't sufficient--and cannot be tested on them. I would even question if they may be asked of adults who do not have a special theological orientation. In its original form which contained 43 questions it was validated on the basis of responses of 67 people, rabbis and theological students, and here too highly varied answers were received. More than 50% answered in the negative questions to which the author expected a positive reply and on the question concerning getting things by praying for them he received only 7% positive responses. As one educator said, "All these questions constitute the subject matter of medieval Jewish philosophy and Jews could never reach an agreement on them; and whatever was accepted by Maimonides was rejected by Joseph Albo"(112).

To illustrate, one may hope that "God protects from harm those who trust in him," but how many of those who do think about such matters are ready to say that they "think" so? We may hope so, but it's a question as old as Job, despite the faith of Abraham--and it may take a lifetime to gain such faith, and one is not irreligious if he is not so sure about it. Are adolescents within any group in Judaism expected to have a firm view about such matters--even if some of these questions are included in confirmation manuals?

Furthermore, the author used this test to make comparisons between orthodox and reform youth, when orthodox schools rarely teach the subject. Orthodox schools are "primarily interested in actual practice, prayer and study, than in the inculcation of theological beliefs" (111). Of course it was validated on adults because the author could not find a truly objective criterion of children's "religiosity."

We have enough difficulty in establishing criteria for evaluating attitudes



which are constantly taught and stressed, without getting involved in testing adolescents in beliefs on which there is no agreement, on which there are numerous interpretations and which are not stressed and often not even taught. The test is unnecessary and invalid for the groups intended.

7. Summary: In summarizing the current status of testing and evaluation in Jewish education, we quote a statement by Edward Nudelman: "While it is true that various small attempts have been made to cope with the problem, there has as yet appeared no truly significant contribution to the measurement of the results of teaching in the Jewish school." (52). This statement was made in 1929 and is, by and large, equally true now. Viewing the situation in 1962, Zalmen Slesinger said: "To the Jewish school the concept of evaluation is, in the main, still foreign and unfamiliar...Progress in Jewish education is utterly impossible in the absence of a comprehensive and effective program of evaluation." (94)

A significant advance in recent years is the establishment of the National Curriculum Research Institute. However, this is still largely a hope rather than a realization, for the NCRI does not have the means to conduct the numerous projects which it must undertake. Its one accomplishment in this area to date is the creation of a standardized achievement test in Jewish history. This counteracts the early period when the Union of American Hebrew Congregations and the Jewish Education Committee of New York issued standard tests in several areas. At present the American Association for Jewish Education in addition to the work of the NCRI, also issues the only other partially validated national test, the Hebrew achievement test, with the Chomsky test also being reviewed for a similar purpose by another organization.



The situation has greatly improved in achievement testing in Bureau affiliated schools, both qualitatively and quantitatively. Twenty of the twenty-seven fully-functioning central agencies for Jewish education conduct achievement tests, many of them creating their own instruments. In most of these communities, however, the testing is mainly in Hebrew, with a maximum of about 20% of the children in these schools tested in all subjects. While most of these instruments meet the criteria of objectivity and practicability, very few conduct an item analysis or maintain other standards of validity and reliability, and in all but one case they make no effort at establishing norms. Only beginnings are noted in testing in the schools of Reform Judaism. In secondary education, except for Hebrew there is also virtually no organized testing. Also, the vast majority of the studies in this area took place a decade and more ago.

In the teaching of Hebrew, the major subject of instruction in the Jewish schools in America, there are several objective, widely-used tests in addition to the two mentioned above. However, there is not even one fully-validated test which can be used nationally without hesitation and with the full realization that it represents an adequate gauge of the success of our children's learning. In Siddur reading there is one satisfactory individual test and the old Nardi group test could be used if revisions were made in it. There are few tests in the content and significance of the Siddur and none is completely satisfactory. The existing tests on holidays and observances are generally not based on specific texts or other specific content, and there is not one which does not have significant omissions or needed changes. In Humash there is virtually no test which can be used in the vast majority of Hebrew Schools in America. In history, in addition to the NCRI test there are several satisfactory objective



tests.

There is little effort in the entire field of Jewish education in creating instruments for evaluating any area other than achievement. In the past two decades there appeared no satisfactory instrument for the testing of Jewish practices, but the earlier Nardi "Home Environment Test" bears review and adjustment. Of the twelve instruments on values and attitudes, only brief questionnaires on attitudes toward the Jewish school are available; others must either be changed or made up to date. The area of specifically Jewish values which is most directly related to the results of the Jewish school learning has no comprehensive and satisfactory testing instrument.

The situation is similar in testing on Hebrew prognosis. Several efforts were made in this area in the past but in recent years there was very little done and at present there is only one partially validated prognosis test in existence, prepared two decades ago, and this needs revision.

In psychological testing we have information of only one community conducting a continuous and comprehensive program, with partial programs in one other community and in several individual schools. In this field there are some well-validated national tests in general education, but the Jewish school simply does not use them. The Jewish school is apparently not convinced of the need for this means of evaluation.

The weakness in this area is further reflected in the absence of professionals devoting themselves exclusively to this work. While two decades and more ago two agencies had full-time directors of testing and experimentation, there is none at the present time, although on a part-time basis many more devote themselves to this work than before.



V. RECOMMENDATIONS

Improvement in testing and evaluation in Jewish education, if it is not to be a slow, haphazard process, must have central overall direction. The instrumentality for this overall direction exists in the National Curriculum Research Institute. This body is to effect noticeable progress in this field by setting standards, intensifying its own work and encouraging and coordinating other efforts. This, of course, requires additional funds and manpower which, it is hoped, the American Association for Jewish Education, its parent body, will be willing and able to provide.

1. a) The NCRI is to convene a conference on testing and evaluation with the help of leaders in this field, in general education, for the purpose of providing initial stimulation and guidance to Bureau directors, supervisors and educational directors of large institutions.

b) It is also to encourage the organizations of Jewish educators: NCJE, EA, NATE, Hebrew teachers and Day School principals to conduct sessions at their coming conferences devoted to the place which testing and evaluation must occupy in their schools. This will succeed only if the NCRI will be able to provide the people to lead these sessions.

2. The NCRI should a) provide guideposts for the creation of standard tests and other instruments in Jewish education and b) farm out the preparation of a number of instruments among the larger Bureaus in the country to be prepared under its guidance. In the immediate future it should consist of the following:



- (1) Parallel tests to its Hebrew Achievement Tests
- (2) Standard tests in Holidays and Observances, Content and Significance of Siddur, Humash Bereshit
- (3) Revised Nardi's Prognosis Test in Hebrew and Group Test in Siddur
- (4) Questionnaire on Jewish practices of students and on Jewish activity of graduates
- (5) Questionnaire on Jewish values and attitudes and revision of questionnaire on attitudes toward Jewish school

These are to be preceded by the establishment of objectives, criteria and subdivisions in each area of study. Following the completion of these instruments it shall administer them in varied communities and establish norms, which shall be revised from time to time.

3. The NCRI shall prepare or direct the preparation of a manual on Testing in the Jewish School- A Guide to Teachers. This should include a) The importance of testing, b) Areas of testing, c) Time Table of testing, d) Available tests and other instruments, e) Criteria for school and teacher-made tests, f) The administration and scoring of tests.

4. It shall also prepare sample sets of available evaluative instruments: achievement tests, questionnaires and check lists on practices and attitudes, psychological tests, check list on textbooks, accreditation forms. These shall be made available to schools and individual teachers.

5. It shall encourage selected communities to conduct testing 3, 6 and 10 years after the termination of studies.

6. Equally important in evaluation is a series of studies on different methods of teaching Hebrew and other subjects of study, the relation of motivation to learning, the influence of the home on children's Jewish education, the validity of I.Q. tests as Hebrew predictors. Here, too, the studies may be conducted by



individual communities but criteria should be carefully prepared and continuous guidance provided to assure validity of the results.

At the same time Bureaus of Jewish Education must, of course, continue or begin to prepare tests on various areas of study and endeavor to enter fields of evaluation not hitherto penetrated. Some of the activities listed above can, and in the absence of national guidance should, be undertaken independently by Bureaus and by national denominational groups, but this will result in varied standards and in unnecessary duplication. National guidance and coordination is indispensable. National guidance and local execution will help provide the standards of evaluation essential in American Jewish education.



BIBLIOGRAPHY

1. Boykin, L.L. "What Is Evaluation" J.Ed. Res. 51:529-34, 1958.
2. Harris, Chester W. Editor, Encyclopedia of Ed. Research, Third Edition, Macmillan, 1960.
3. Justman, J. "Evaluation in the Modern School," Jewish Education 29:1,10-16, 24, 1958.
4. Smith, E. R., Tyler, R. W. and the Evaluation Staff, Appraising and Recording Student Progress, Harper and Brothers, 1942.
5. Wrightstone, J. Wayne, Justman, Joseph, Robbins, Irving, Evaluation in Modern Education, American Book Co., 1958.
6. Wrightstone, J. Wayne. Appraisal of Newer Elementary School Practices, T. C. ' 38, Pp. 152-53.
7. Boykin, Leander L. "Let's Eliminate the Confusion:What Is Evaluation" Ed. Ad. and Sup. 43: 115-21, 1957.
8. Orate, Pedro T., "Evaluating Evaluation," J.Ed.Res. 33:9, 641-661, 1940.
9. Monroe, Paul, A Cyclopedia of Education, Macmillan, 1913.
10. Good, C. V., Editor, Dictionary of Education, McGraw-Hill, 1959.
11. Scates, Douglas E., "Fifty Years of Objective Measurement and Research in Education" J. Ed. Res. 41: 241-64 D'47.
12. a) Hosic, J. F. "The Problem of Evaluation," Ed. Method, 15: 448, May '36.  
b) Heffernan, Helen, "Evaluation More Than Testing," NEA Journal 4/58, pp. 227-29.
13. Monroe, Webber S. "Ed. Measurement in 1920 and 1945" J. Ed. Res. 38: 334-40, 1945.
14. Leavitt, Howard B. "The Dichotomy Between Ends and Means in American Education" Journal of Education 141: 14-16, 1958.
15. Findley, Warren G., "Ed. Evaluation-Record Developments," Social Ed., 14: 206-10, May '50.
16. Russell, David H, "What Does Research Say About Self-Evaluation," J.Ed.Res., 46: 561-73, 1953.
17. Strang, Ruth, "The Interview," Review of Ed. Res. 9:498-501, D'39.
18. Traxler, Arthur E., Techniques of Guidance.
19. Hubbard, Frank W., "Questionnaires" Review of Ed. Res., 9:502-507, D'39.
20. Traxler, Arthur E., "A Cumulative Record Form for the Elementary School" Elementary School Journal, 40:45-54, 5'39.



21. a) Olson, Willard C. "The Case Method" Rev. Ed. Res. 9:483-90, D'39.  
b) Traxler, Arthur E. "Case Study Procedures in Guidance," School Review 46:602-10 O'38.
22. Jersild, A. T. and Maigs, M.F. "Direct Observation As Research Method" Rev. Ed. Res. 9:472-82 D'39.
23. Wrightstone, J. Wayne, "Constructing An Observational Technique," T. C. Record 37: 19 O'35.
24. Michaelis, John U. and Howard, Charles "Current Practices in Evaluation in City School Systems in Calif." J. Ed. Res. 12/49.
25. Partzer, Clyde A., Wright, E. Wayne and Clark, Selby G. "Questions Concerning the Interview As A Research Technique." J. Ed. Res. 51:215-21 N'57.
26. Ryans, David G. "Are Ed. Research Offices Conducting Research?" J. Ed. Res. 51: 173-83 N'57.
27. Sears, Jesse B. "School and Community Surveys: Rev. Ed. Res. 9:508-513 O'39.
28. Diederich, Paul B. "Design for a Comprehensive Evaluation Program" School Review 58:225-32 Ap'50.
29. Freeman, Frank S., Theory and Practice of Psychological Testing, Rev. Ed., Holt, 1955.
30. Whitestone, J. Wayne, "Testing and Evaluation" NEA Journal 47:221-23 Ap'58.
31. Rivlin, Harry N. Encyclopedia of Modern Education, 1943.
32. Mayhew, Lewis B. "The Critical Incident Technique in Educational Evaluation," J. Ed. Res. 49:591-98, 1956.
33. Linquist, E. F., Editor, Educational Measurement, American Council on Ed., 1950.
34. "Technical Recommendation for Psychological Tests and Diagnostic Techniques," Supplement to Psychological Bulletin 51: 1-38, Mar'54.
35. Thorndike, Robert L. and Hagen, Elizabeth, Measurement and Evaluation in Psychology and Education, Second Ed. John Wiley and Sons, 1961.
36. a) ETS Builds a Test  
b) Katz, Martin, "Selecting An Achievement Test" ETS, Series #3.
37. Diederich, Paul B., "Short-Cut Statistics for Teacher-Made Tests," ETS Series #5, 1960.



38. Smith, G. Milton A Simplified Guide to Statistics, Rinehard & Co, 1950.
39. a) Committee on Test Standards, American Educational Research Association and National Council on Measurements Used in Education, Technical Recommendations for Achievement Tests, NEA, 1955.
- b) American Psychological Association, Technical Recommendations for Psychological Tests, 1954.
40. Slesinger, Zalman "Basic Considerations in a Program of Evaluation", The Pedagogic Reporter, Vol. IX, no. 1, 9/57, P.10.
41. Maller, J. B., "Testing the Knowledge of Jewish History", UAHC, Cincinnati, 1932.  
See also Jewish Education, Volume II; 2,3 6/30 and 10/30 for selected sections.
42. Dushkin, Alexander M. and Engelman, Uriah Z., Jewish Education in the United States, Volume I, AAJE, 1959.
43. Kaplan, Mordecai M. and Cronson, Bernard, "The First Community Study of Jewish Education in New York City, 1909", Jewish Education 20;3, Summer 1949.
44. Benderly, Dr. S., "Aims and Activities of the Bureau of Jewish Education of the Jewish Community (Kehillah) of New York 1912", JE Volume XX: 3, Summer, 1949, Pp. 92-109.
45. Garfiel, Evelyn N. "Hebrew Prognosis Test," Journal of Semitic Languages and Literatures, Vol. XLV:1, October, 1928.
46. Vanek, Zdanek, "Testing in American Jewish Education", JE 25,3 Winter 54-55, 36-45 (condensation of thesis).
47. a) Hurwich, Louis, "Jewish Education in Boston", JE, 26:3, Spring 1956, pp.22-34.  
b) JE, Volume I:1, 1/29, P.62.  
c) Hurwich, Louis, "Standard Examinations in Greater Boston Hebrew Schools", JE, Vol. I: 2, 5/29, pp. 14-18.
48. Bugatch, Simon "Supervision Through Testing in Baltimore", JE, Vol.21:2, Spring 1950-pp. 33-39.
49. Gannes, Abraham P., Central Community Agencies for Jewish Education, Dropsie College, 1954, 242 pp.
50. Fishman, Joshua A., "Evaluation of Results in Current American Jewish Education"- JE, Vol.24;3, Spring '54, pp. 22-28.



51. Survey of Jewish Education in Greater New York, 1951-52, Qualitative and Quantitative Studies, JEC, 1959
52. Nudelman, Edward A., Review of book "Objective Tests" by Jacob S. Orleans and Glenn A. Sealy-Woll- JE 1:2, 5/29, pp. 137-38.
53. Maller, Julius B., "Needed Investigations in Jewish Education", JE 1:2, 5/29, pp. 97-109.
54. Spotts, Leon H., Review of "Evaluation and the Elementary Curriculum" by Harold G. Shane and E. T. McSwain, JE, 29:3, Spring '59.
55. Maller, Julius B., "A Standardized Hebrew Prognosis Test", JE 3:3, 10-12/31
56. Golub, Jacob S., Review of "Jewish History Achievement Tests" by Dr. Julius B. Maller and Rabbi Jacob Pollak, JE 2:2, 6/30 pp. 106-7
57. Nardi, Noah, "Group Test for Siddur Reading", JE, 18:1, 11/46, pp. 49-51
58. Nudelman, Edward A., "Index for Testing Reading Facility", JE 1:2, 5/29, pp. 121-23.
59. Bugatch, Simon, letter in JE 29:3, Spring '59, p. 60
60. Benathen, Morris, "Basic Vocabulary and Language Forms of the Siddur", JE, 29:2, Winter 1959
61. Hartstein, Jacob I., Symposium on "Common Elements in Jewish Education" Traditionalist View, JE, Vol. XVII:2, 2/46.
62. *Handwritten note in Hebrew:* "הנכנסים ללימודי תורה, חייבים להיות מוכנים ללמוד, וללמוד, וללמוד."
63. a) Freehof, Solomon B., "Guiding Principles of the Commission on Jewish Education" in An Outline of the Curriculum for the Jewish Religious School, UAHC.  
 b) Ibid, pp. 11-12.
64. United Synagogue Commission on Jewish Education, Objectives and Standards for the Congregational School, 1958.
65. Jung, Leo and Joseph Kaminetsky, Editors, A Model Program for the Talmud Torah, UOJCA 1942.
66. Dushkin, Alexander M., "Common Elements in Jewish Education", JE XVII:1, 11/45.
67. Perlmutter, Elsie K., "The Role of the Case Worker in the Jewish School", JE, 28:2, Winter 57-58, pp. 38-41.
68. Gelbart, Frances Spitzer, "An Experiment in Case Work Service in a Pre School Program" JE, 28:2, Winter, 57-58, pp. 42-48.
69. Goodnick, Benjamin, "Evaluating Pupil Growth", Pedagogic Reporter, VIII:1, 9/56.



70. Golub, Jacob S., Berkson, I.B., Edidin, Ben M., and Dinin, Samuel.,  
"Evaluating Jewish School Curricula," JE XI:3, 1/40.
71. Outline of Curriculum for Congregational Hebrew Schools, JEA, Essex County,
72. Bannett, Alan D., "Curriculum and Materials", Educational Research Survey  
No. 5, UAHC, 1962.
73. Jewish Parent, 13:1, 6/61.
74. JEC Bulletins, 11/61, 5/62.
75. Gross, Morris, "Psychological Services for Yeshivoth", JE, 30:3, Spring 1960.
76. Lerner, Benjamin, "The Psychologists Role in the Day School", Jewish Parent,  
14:2, 10/62.
77. Survey of Jewish Education in Greater New York 1951-52-Qualitative Studies.
78. Jewish Teacher, Vol. VII:4, 6;39, IX:1, 11/40, IX:2, 1/41, IX:3, 4/41.
79. Educational Notes, Los Angeles, 1/57.
80. Shapiro, Murray, "Evaluating Teacher-Constructed Social Studies Units",  
Jewish Teacher, Vol. 30:3, 2/62.
81. Hartstein, Mrs. J.I., article in Jewish Parent, 11:3, 12/15/59.
82. Ettinger, Celia, A Study of the Portland (Oregon) Hebrew School, 9/62, in  
partial fulfillment of the requirement for the degree Master of Arts in  
Teaching, Reed College.
83. Dunkel, Harold B., Second Language Learning, Ginn & Co., Chicago, 1948.
84. Cole, Robert D., Modern Foreign Languages and Their Teaching, D.Appleton-  
Century Co., 1937.
85. Nardi, Noah, "Studies in Intelligence of Jewish Children", JE, 19:3, 15, 19:3,  
Sum 1948  
Summer 1948, pp. 41-51.
86. Nardi, Noah, "A Test to Measure Aptitude in the Hebrew Language", Journal of  
Educational Psychology, Vol. 38:3, 3/47, pp. 167-76.
87. Ross, C.C. and Hooks, U.T., "How Shall We Predict Achievement," J.Ed.Res.,  
XXII:3, 10/30, pp. 184-96.
88. Bovee, A. G. and Froehlich, J.G., "Some Observations on Relation Between  
Mental Ability and Achievement in French", School Review, LIII:9, 11/45,  
pp. 534-7.



89. Spoerl, Dorothy T., "A Study of Some Possible Factors Involved in Foreign Language Learning," Modern Language Journal, XXIII:6, 3/39, pp. 428-31.
90. Von Wittich, Barbara., "Prediction of Success in Foreign Language Study," Modern Language Journal, XLVI:5, 5/62, pp. 208-12.
91. Tallent, E.R.E., "Three Coefficients of Correlation That Concern Modern Foreign Languages", Modern Language Journal, XXII:8, 5/38, pp. 591-4.
92. Kohn, Eugene, "Character Training in the Jewish School," JE, III:1, Jan. 1931, pp. 20-28.
93. Amsel, Berenice, "Testing or Overtesting" Jewish Parent XIV:3, January 1963.
94. Slesinger, Zalman, "Evaluation, the Key to Progress in Jewish Education," Pedagogic Reporter, XIII:3, January 1962.
95. Chomsky, William, Teaching Hebrew, JEC 1956.
96. Pollock, George, Review of "Test on the Fundamentals of Hebrew," Synagogue School, XX:1, 12/61.
97. Poisson, Joseph A., "The Role of Evaluation in the Hebrew School System", Synagogue School, XX:2, 6/62.
98. Ruffman, Louis L., "School Accreditation in Greater New York," JE 32:2, Winter '62.
99. Fishman, Joshua A., "Educational Evaluation in the Context of Minority Group Dynamics," JE, XXIX;1, Fall 1958.
100. Spotts, Leon H., Review of "Sociometry in the Classroom," by Norman Nroulund, JE 31:1, Fall '61, P. 62.
101. Brav, Stanley R., "Attitudes Toward Jewish Education, A Sampling of Reform Jewish Education", JE, Vol. 24:3, Spring '54, pp. 35, 60.
102. Eisenberg, A. and Warkow, Seymour, "Continuity of Higher Hebrew Study and Jewish Home Environment, A Study of 592 Students", JE, 26:3, Spring '56, pp. 42-50.
103. Zeligs, Rose, "Racial Attitudes of Jewish Children", JE, 9:3, Oct.-Dec., 1937, pp. 148-52.
104. Blumenfield, Samuel M., "A Study of the Correlation Between Elementary Jewish Education and Interest In Jewish Life and Problems", JE, 9:3, Oct.-Dec., 1937, pp. 143-47.



105. Baylor, Ruth M., "How Well Do We Teach, The Evaluation of the Development of Character Through Bible Stories," Jewish Teacher, XIX:1, 11/50, pp.19-25.
106. Nathan, Marvin, "The Attitude of the Jewish College Student Toward Judaism", JE, IV:1, 1/32, pp. 48-53.
107. Dinin, Samuel, "The Curriculum of the Jewish School," American Jewish Year Book, 1962, pp.214-25.
108. Chomsky, William, "Experience With A Testing Program in Hebrew", JE, XIV:3, Jan.-March, 1943, pp. 134-7.
109. Nardi, Noah, "Studies in Jewish Home Environment," JE, 17:2, 1946, pp.44-49.
110. Franzblau, Abraham N., Religious Beliefs and Character Among Jewish Adolescents, Teachers College, 1934.
111. Dinin, Samuel, Review of "Religious Beliefs and Character Among Jewish Adolescents" by Abraham N. Franzblau, JE, VII:2, Spring, 1935.
112. Soviv, Aaron, "Self-Acceptance of Jewishness by Young Jewish People," JE, XVI:1, Summer, 1955, pp. 22-31.